

GenCore version 5.1.4-05-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 28, 2003, 13:21:54 ; Search time 81 Seconds
(without alignments)
1579.018 Million cell updates/sec

Title: US-09-769 699-2

Perfect score: 6294

Sequence: 1 MENTUKITVTGTLGIVYA DELFTLSGIPKIKHNTMEM 1203

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 408470 seqs, 13525046 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: A.Geneseq_101002:*
2: /SID2/gcqdala/geneseq/genesep-emb1/AA1980.DAT:*
3: /SID2/gcqdala/geneseq/genesep-emb1/AA1981.DAT:*
4: /SID2/gcqdala/geneseq/genesep-emb1/AA1982.DAT:*
5: /SID2/gcqdala/geneseq/genesep-emb1/AA1983.DAT:*
6: /SID2/gcqdala/geneseq/genesep-emb1/AA1984.DAT:*
7: /SID2/gcqdala/geneseq/genesep-emb1/AA1985.DAT:*
8: /SID2/gcqdala/geneseq/genesep-emb1/AA1986.DAT:*
9: /SID2/gcqdala/geneseq/genesep-emb1/AA1987.DAT:*
10: /SID2/gcqdala/geneseq/genesep-emb1/AA1988.DAT:*
11: /SID2/gcqdala/geneseq/genesep-emb1/AA1989.DAT:*
12: /SID2/gcqdala/geneseq/genesep-emb1/AA1990.DAT:*
13: /SID2/gcqdala/geneseq/genesep-emb1/AA1991.DAT:*
14: /SID2/gcqdala/geneseq/genesep-emb1/AA1992.DAT:*
15: /SID2/gcqdala/geneseq/genesep-emb1/AA1993.DAT:*
16: /SID2/gcqdala/geneseq/genesep-emb1/AA1994.DAT:*
17: /SID2/gcqdala/geneseq/genesep-emb1/AA1995.DAT:*
18: /SID2/gcqdala/geneseq/genesep-emb1/AA1996.DAT:*
19: /SID2/gcqdala/geneseq/genesep-emb1/AA1997.DAT:*
20: /SID2/gcqdala/geneseq/genesep-emb1/AA1998.DAT:*
21: /SID2/gcqdala/geneseq/genesep-emb1/AA1999.DAT:*
22: /SID2/gcqdala/geneseq/genesep-emb1/AA2000.DAT:*
23: /SID2/gcqdala/geneseq/genesep-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1203	22	AA1980.DAT
2	3016.5	47.9	1196	19	AA1981.DAT
3	1566.5	24.9	623	19	AA1982.DAT
4	1380.5	21.9	543	19	AA1983.DAT
5	620	9.9	1134	21	AA1984.DAT
6	183	2.9	35	20	AA1985.DAT
7	183	2.9	35	22	AA1986.DAT
8	156.5	2.5	1841	21	AA1987.DAT
9	123.5	2.0	3210	22	AA1988.DAT
10	123	2.0	1074	20	AA1989.DAT

11	123	2.0	1074	20	AA1990.DAT
12	123	2.0	1074	23	AA1991.DAT
13	123	2.0	1074	23	AA1992.DAT
14	122.5	1.9	901	22	AA1993.DAT
15	122.5	1.9	2387	21	AA1994.DAT
16	121	1.9	1069	21	AA1995.DAT
17	120.5	1.9	2597	21	AA1996.DAT
18	120.5	1.9	2597	23	AA1997.DAT
19	118.5	1.9	1076	22	AA1998.DAT
20	118	1.9	1784	22	AA1999.DAT
21	115	1.8	1150	14	AA1980.DAT
22	114.5	1.8	1496	20	AA1981.DAT
23	114.5	1.8	1496	21	AA1982.DAT
24	114.5	1.8	1498	22	AA1983.DAT
25	114	1.8	774	7	AA1984.DAT
26	111	1.8	4630	18	AA1985.DAT
27	111	1.8	4630	21	AA1986.DAT
28	110.5	1.8	1532	22	AA1987.DAT
29	110.5	1.8	3070	22	AA1988.DAT
30	109	1.7	11096	22	AA1989.DAT
31	108.5	1.7	1451	22	AA1990.DAT
32	108	1.7	2596	22	AA1991.DAT
33	106.5	1.7	794	16	AA1980.DAT
34	106.5	1.7	3298	23	AA1981.DAT
35	106.5	1.7	3567	14	AA1982.DAT
36	105	1.7	774	18	AA1983.DAT
37	104.5	1.7	1471	19	AA1984.DAT
38	104.5	1.7	1515	23	AA1985.DAT
39	104.5	1.7	3530	22	AA1986.DAT
40	104.5	1.7	774	13	AA1987.DAT
41	104	1.7	774	16	AA1988.DAT
42	104	1.7	821	13	AA1989.DAT
43	103.5	1.6	1750	22	AA1990.DAT
44	103	1.6	774	14	AA1991.DAT
45	103	1.6	774	14	AA1992.DAT

ALIGNMENTS

RESULT 1
ID AAM52633 standard; Protein: 1203 AA
XX AC AAM52633:
XX DT 18-FEB-2002 (first entry)
XX DE Varicella-zoster virus (VZV) ORF29p protein
XX KM ORF29p: open reading frame 29p protein; Varicella-zoster virus; VZV:
XX KM drug delivery; gene delivery; fusion protein; protein secretion.
XX KW Human herpesvirus 3.
XX OS
XX PN W0200154709-A1.
XX PD 02-AUG-2001.
XX PE 25-JAN-2001; 2001WO-US02500.
XX PP 25-JAN-2000; 2000US-0177901
XX PA (HVC) UNIV COLUMBIA NEW YORK.
XX PI Silverstein S, Annunziato P, Gershon A, Lungu O:
XX XX WPI, 2001-488744/53.
XX DR N-PSDB; ABA02175.
XX PT Composition of matter for delivering desired compounds into eukaryotic
XX PT cell, comprises Varicella-zoster virus open reading frame 29p protein
XX PT either bound to an agent or operably affixed to lipid soluble group

XX Claim 1: Fig 6: 68pp: English.

XX This sequence represents the Varicella-Zoster virus (VZV) open
 XX reading frame 29p (ORF29p) protein. Although ORF29p is the major DNA
 XX binding protein of VZV, the invention is based on the discovery that the
 XX VZV ORF29p protein can readily enter and exit eukaryotic cells.
 XX Accordingly, the invention relates to a composition for delivery of an
 XX agent into a eukaryotic cell, comprising the VZV ORF29p protein bound to
 XX the agent to be delivered. The agent delivered may be a polypeptide, a
 XX polynucleotide, or an organic compound. The invention also relates to
 XX a composition in which the ORF29p protein is linked to a lipid soluble
 XX group that permits the protein to be anchored to a lipid membrane, and
 XX lipid vesicles containing this composition. The invention additionally
 XX relates to the use of the ORF29p protein as a facilitator of secretion of
 XX a desired protein, in which the desired protein and the ORF29p protein
 XX are recombinantly expressed as a fusion protein. The invention further
 XX encompasses nucleic acids encoding the VZV ORF29p protein and their use
 XX in detection of VZV ORF29p nucleic acids, and a monoclonal antibody
 XX against the ORF29p protein. Compositions of the invention are useful for
 XX the delivery of a prophylactic or therapeutic agent to a eukaryotic cell,
 XX particularly a human cell. Examples of agents that can be delivered to a
 XX cell include proteins such as insulin, factor VIII, factor IX, and
 XX proteases; polynucleotides (e.g., for use in gene therapy); and organic
 XX compounds such as vitamins and a wide variety of pharmaceuticals (e.g.,
 XX antineoplastic, and labeled or immunosuppressive drugs).

XX Sequence 1203 AA:

Query Match: 100.0%; Score 6,294; DB 22; Length 1203;
 Post Local Similarity 100.0%; Pred. No. 0;
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 1 MENTQKTVTPVSPGAYVACREVDJDEEISPLAARSTDSIALPLMNLVEKTFIS 60
 DB 1 MENTQKTVTPVSPGAYVACREVDJDEEISPLAARSTDSIALPLMNLVEKTFIS 60
 UY 61 SLAVVSGAATGTLGAGGTLKLTSHFYPSVYFPGGKIVLPSSAAPNLTRACNAAREP 120
 DB 61 SLAVVSGAATGTLGAGGTLKLTSHFYPSVYFPGGKIVLPSSAAPNLTRACNAAREP 120
 UY 121 GPSKVGPPVKAIVETTCATCTRLCEPENTLLYLVATLKEAFMGNVFLHYGLDI 180
 DB 121 GPSKVGPPVKAIVETTCATCTRLCEPENTLLYLVATLKEAFMGNVFLHYGLDI 180
 UY 181 VHNHGDVRIHLEFVQLEFMPVNNKLVPPDPNTHRSIGCFVYVPPYNTGLIHNC 240
 DB 181 VHNHGDVRIHLEFVQLEFMPVNNKLVPPDPNTHRSIGCFVYVPPYNTGLIHNC 240
 UY 241 VIAPMAVALKNNVNTAVAGCAHLAFDENIEGAVLPDITTYEFGSSSGTTARCARRN 300
 DB 241 VIAPMAVALKNNVNTAVAGCAHLAFDENIEGAVLPDITTYEFGSSSGTTARCARRN 300
 UY 401 IWNSTKSPSPSPSEFEERASIMAAVTALAEVFNFTGIEEPTDTRKEMPEFGMETLP 360
 DB 401 IWNSTKSPSPSPSEFEERASIMAAVTALAEVFNFTGIEEPTDTRKEMPEFGMETLP 360
 UY 461 KLMALASTYARAVAGVIGAVVSPNSALYTFEVDNSGMLFAKNGGPPSPNNRYQFAGPHL 420
 DB 461 KLMALASTYARAVAGVIGAVVSPNSALYTFEVDNSGMLFAKNGGPPSPNNRYQFAGPHL 420
 UY 461 KLMALASTYARAVAGVIGAVVSPNSALYTFEVDNSGMLFAKNGGPPSPNNRYQFAGPHL 420
 DB 461 KLMALASTYARAVAGVIGAVVSPNSALYTFEVDNSGMLFAKNGGPPSPNNRYQFAGPHL 420
 UY 421 AANPOTDRKHIVLSSSGTSSNTEFSDVLAICGGAFLALFLYLERGCGAFTGGH 480
 DB 421 AANPOTDRKHIVLSSSGTSSNTEFSDVLAICGGAFLALFLYLERGCGAFTGGH 480
 UY 441 GHALKYVTGTFDSEIFGSLCEKHTPRVCAHTTVHRLRQMPREGQATROPIGVEGTMSQ 540
 DB 441 GHALKYVTGTFDSEIFGSLCEKHTPRVCAHTTVHRLRQMPREGQATROPIGVEGTMSQ 540
 UY 481 GHALKYVTGTFDSEIFGSLCEKHTPRVCAHTTVHRLRQMPREGQATROPIGVEGTMSQ 540
 DB 481 GHALKYVTGTFDSEIFGSLCEKHTPRVCAHTTVHRLRQMPREGQATROPIGVEGTMSQ 540
 UY 541 VSDVTPLEGVAVYLLIKRKQDQTEAAKATIMQDYRATLERFLDEQERLDRGACSSS 600
 DB 541 VSDVTPLEGVAVYLLIKRKQDQTEAAKATIMQDYRATLERFLDEQERLDRGACSSS 600

UY 601 GLSSVIVDHPPTPRILLDTPARIEGTTOPMKVLEVETPKVIRKGLSEATHSMALTPDPY 660
 DB 601 GLSSVIVDHPPTPRILLDTPARIEGTTOPMKVLEVETPKVIRKGLSEATHSMALTPDPY 660
 UY 661 SGAPCPITNLTAKRTHLAVVDLALSYCHCVYGVGVGCAKPKKNOVQVYLRKKRVDFPNC 720
 DB 661 SGAPCPITNLTAKRTHLAVVDLALSYCHCVYGVGVGCAKPKKNOVQVYLRKKRVDFPNC 720
 UY 721 GFISTSTITTSISGVSANPLISGVAIPAGKTFHGLAVASVAVIKDVRKNNVNSGN 780
 DB 721 GFISTSTITTSISGVSANPLISGVAIPAGKTFHGLAVASVAVIKDVRKNNVNSGN 780
 UY 781 CTNLSAARAKVLGLASAVYRUEKRVDMGLGALFLLKDFHGLFPGMPNSKSPNPOW 840
 DB 781 CTNLSAARAKVLGLASAVYRUEKRVDMGLGALFLLKDFHGLFPGMPNSKSPNPOW 840
 UY 841 FWTLLQKNOQMPARKTTHEETTTAAVKRFDEYAAINPILNPPGICGELAGPTMANILK 900
 DB 841 FWTLLQKNOQMPARKTTHEETTTAAVKRFDEYAAINPILNPPGICGELAGPTMANILK 900
 UY 901 YCHSGLINTLSTITGARRKPPSSVLHWIKKDVSAADIFEOAKALLKTEKNEPELW 960
 DB 901 YCHSGLINTLSTITGARRKPPSSVLHWIKKDVSAADIFEOAKALLKTEKNEPELW 960
 UY 961 TIAFTSTHLVKAAMQKPPVVLITISKYHSAANNEVFAGNMSSLNGKKNVPLETTFD 1020
 DB 961 TIAFTSTHLVKAAMQKPPVVLITISKYHSAANNEVFAGNMSSLNGKKNVPLETTFD 1020
 UY 1021 KTRRFITACPRGCFICVYTPSPSSNNRFTTISQVKGILIVSGAMVGLIATVYVRAVGAR 1080
 DB 1021 KTRRFITACPRGCFICVYTPSPSSNNRFTTISQVKGILIVSGAMVGLIATVYVRAVGAR 1080
 UY 1081 AQHMAFDDMLSLTDDDEFLANDLELADQILQTLFTWTVAGALAAVAILDEKTTAGNGCT 1140
 DB 1081 AQHMAFDDMLSLTDDDEFLANDLELADQILQTLFTWTVAGALAAVAILDEKTTAGNGCT 1140
 UY 1141 PTMLAFNDSGCESSHDTSVNLISGNSISGTVPGIKRPPEDDELFDLSGTP1KKNIT 1200
 DB 1141 PTMLAFNDSGCESSHDTSVNLISGNSISGTVPGIKRPPEDDELFDLSGTP1KKNIT 1200
 UY 1201 MEM 1203
 DB 1201 MEM 1203

RESULT 2
 ID AAW72105 standard; Protein; 1196 AA.
 AC AAW72105;
 DE 18-DHC-1998 (first entry)
 DE HSV-2 strain SD5 Contig ID 12 ORF#2 protein.
 KW HSV-2 strain SD5; immunological response inducer; can. therapy;
 KW antiviral identification; viral protein inhibitor.
 OS Herpes simplex virus type 2.
 PN W09820016-A1.
 PD 14-MAY-1998.
 PP 31-OCT-1997; 97WO-US24010.
 PR 09-JUN-1997; 97NS-0049018
 PR 04-NOV-1996; 96US-0030279.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
 XX Esser KM, Leary JT;

Claim 10; Page 123; 748bp; English.

XX This sequence represents a Herpes simplex virus type 2 (HSV 2) protein
XX sequence of the invention. This sequence was isolated from a HSV-2 strain
XX SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

XX Based on homology, this sequence is a major DNA-binding protein.

XX The proteins can be used for the treatment or prevention of disease, to
XX induce an immunological response in a mammal or to identify inhibitors,
XX activators or novel antivirals. Antagonists of the proteins can be used
XX to inhibit a viral polypeptide. The DNA sequence or a vector containing
XX it can also be used to induce an immunological response in a mammal.

XX Sequence 624 AA:

Query Match 24.9%; Score 1566.5; DB 19; Length 623;

Best Local Similarity 50.0%; Pred. No. 2,90-145;

Matches 113; Conservative 100; Mismatches 192; Indels 21; Gaps 9;

4 TUKTYVTGGLAYYY ACRYEDLLEELFLAANSTDSLLALMLKNTYEKFTSSL 62

7 TTTTAVKPHGCMGVYKATTAAGLEL-LSLSASGSAVAVAVLIVOLTFESGFEAVV 64

65 AAVSGARTGACACILIKLITSHYFVYFHOCKHVLSSAALNIFACMAARHRCF 122

65 AAVSGRTTGLATAVSKIMPSHSPSYVTHGPHLAPSTQANLTRLEBARHRCF 124

123 SPQGPVPCAVETPAALCTHCAEPENTIVYVVAIPKFAVPCVNAHVCGIPVH 182

125 SHYAPRKHCKHETGDMALCDDIDHALLIVITFCRFAVGLSNTECHLQMDKVT 184

184 ENHIVITETITVOLEHMDVAKLVDFEFTIHSTGRGTVPFTFYTGIFLHHPVI 242

185 TGDVAVHPIVYVILMPHDSFVIALPYNHRSICFNFTLCTFPNNIARLFEAVV 244

243 AFMAVALKRVNVTAAKCAHILAFENHCAVILPDDIYTFYSSSSSTTARCARNDV 302

245 GFAVALKAPVIVAVAAHILAFLENHCAVILPDDIYTFYSSSSSTTARCARNDV 298

101 NSTKSPISGCEKRLASIMADTALAAVIFNCTIYETPTDICEKMPICMECHLRL 362

209 DACNKGACGCEPMLASVMAADALALSLVSMVPIPIPIPIPIPIPIPIPIPI 357

363 NALISYTRAVAVICAGVESPNSALYTEVEDSMTKADGSPSENFYGFAPSHLNA 422

398 GAVVATYLAKAANTVCAWYPTNSALHILEVDIAGADIRKDKK-PSYKRPFLVPTHYAA 416

423 NVQDTRDCHVLSSOSTGSSNIFSVYDALICGQCAPILLARLFLYERQDACA 475

417 NPQDREHIVAVYERHPTAPLVGTO-EFAGEHILAMLYGFSPLALLAKMLEYERQDGV 475

476 FEGCHG DALVYVGLIHSEIHCSCGKHTRVCAHTTVHRIKOMPHPGCAATQPIGV 534

476 TVCKQEMVVERVAVNSQTDIVKNCICTFETRHACAHITLMIRAKHPKASAKARACIGF 535

545 GTMNVQVSNPCTFNVAAYLILKPKQCTEAKATMCTEYATLEPFLDEQERLLDPG 594

546 GIMNSAVSNPCTFNVAAYLILKPKQCTEAKATMCTEYATLEPFLDEQERLLDPG 594

595 APVSSGSLSSVAVVAGTTRKRIIDTK 620

595 VPTAARLETTITGREALITVANNIK 620

RESULT 4

10 AAV72196 standard; Protein: 543 AA.

AAV72196:

13 JAN 1999 (first entry)

10 HSV 2 strain SB5 Contig ID 15 ORF#31 protein.

KM HSV 2 strain SB5, immunological response induction; therapy;
KM antiviral identification; viral protein inhibitor.

OS Herpes simplex virus type 2.

PN W09820016-AL.

PD 14-MAY-1998.

PF 31-OCT-1997; 97WO-US20016.

XX 09-JUN-1997; 9/US-0049018.

PR 04-NOV-1998; 96US-0040279.

XX (SMK) SMITHKLINE BEECHAM CORP.

PI Chan JY, Dabrowski AM, Delvecchio AM, Dillon SH;

PI Esser KM, Leary JT;

DR WPI: 1998-286847/25.

XX N-PSDH; AAV62176.

PS Claim 10, Page 122-123; 748bp, English.

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein

CC sequence of the invention. This sequence was isolated from a HSV-2 strain

CC SB5 (deposited as ATCC VR-2546) DNA fragment designated contig ID 15.

CC Based on homology, this sequence is a major DNA-binding protein.

CC The proteins can be used for the treatment or prevention of disease, to

CC induce an immunological response in a mammal or to identify inhibitors,

CC activators or novel antivirals. Antagonists of the proteins can be used

CC to inhibit a viral polypeptide. The DNA sequence or a vector containing

CC it can also be used to induce an immunological response in a mammal.

XX Sequence 543 AA:

Query Match 21.9%; Score 1380.5; DB 19; Length 543;

Best Local Similarity 48.7%; Pred. No. 3,48-125;

Matches 271; Conservative 101; Mismatches 165; Indels 19; Gaps 7;

QY 653 MALTRVPSGACPIITNPLVKKTHLAVVDIALSYCHCYVQGVGNRPKQPVVLR 712

DB 1 MSLSIDPPTGCGCPLOLLARRSNLAAYVDIALSOCHGVAGQSVGNRPKQPVVLR 60

QY 713 RVDVFNKGFTSTRTTTLSEG-PVSAPNPTVQDPAPGPTFGDIAVSEVETDIPV 771

DB 61 RMDLPFNNQPLSAKTLITVALSDEGALCALSLTACOTAPAKSSPQGVAVVITGPKRLRV 120

QY 772 KRVVPSGNCINI-SFAARALVGLASAYOKKKVMDLHGALCFLLKQPHGLPROMP 831

DB 121 KSRVLFAGSASANAFAKRVASLQSAVOKPEKRVIDILGPGELKKPHAVIFNGKPE 180

QY 832 NKSVPNQVFWLLORNQNPADKLTHERITTAARKRTFERHAALNFNLPITGHLAQ 891

DB 181 GSNQNPQVFWLADQNPADKLSHIDITLAFIKRPSLDGALNFNLAANNVSEIAM 240

QY 892 EYMANILKYQDHSYLLINTLSITGARPRDPSVLIMIKDVTSAADITOKALLE 951

DB 241 YMANQILKRYCHSYIFINTLFAVLAGSKRPYGVAAAMAVQ---GGAGILAGAKALMD 297

QY 952 KVENIPELMTTATSTHLVRAAMNPPVAVGISISKYGAAGNNVPOAGWSSDNGK 1011

DB 298 SLDAHGAWTSKFAACNLDRPYMAARPMVVLGLSTISKYYGMAQNDVDFQAGWASLIGK 357

QY 1012 NVCPILFTRDRIRPFIACGRGFGICVNPQSSGNRTTSLDVGRIIVSGAMVQIALYA 1071

DB 358 NCPILFTRDRIRPFIACGRGFGICVNPQSSGNRTTSLDVGRIIVSGAMVQIALYA 1071

QY 1072 TVVRAVGARQHMARDQWLSTDEFLARDELHIDQIQTLETPTVEGAL---EAVK 1127


```

DB 418 AAVKSLGPKHIOUQJEDWALILEPVLSEPMKPTTRALERGHGESTDAALVAHAHA 477
QY 1128 ILDEKTTATGSETPINLAENFSCSESHDTSTNVLANISSNLSKSVPLKKEPDEDELF 1187
DB 478 IYSUJGAAAGH-----VNNKQKPGHJHJHHAASFGSLAA--AGAAQVAKKAKFHJDDPP 528
QY 1188 DLGILPKHGNITMGM 1203
DB 529 G-EGPPEKK-DLTLDM 542

RESULT 5
AAB53126
ID AAB53126 standard: Protein: 1132 AA.
AC AAB53126:
DP 28-FEB-2001 (first entry)

Macaca mulatta rhadinovirus 17577 RRV ORF6 protein SEQ ID NO:6
XX
XX Macaca mulatta rhadinovirus 17577: RRV: rhesus macaque rhadinovirus;
KM genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KM IL-6, macrophage inflammatory protein; MIP; diagnosis, vaccine,
KM cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma,
KM lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KM splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX
XX Macaca mulatta rhadinovirus 17577.
XX
XX WO200028040-A2.
XX
XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26260.
XX
XX 06-NOV-1998; 98US-0107507.
XX
XX 20-NOV-1998; 98US-0109409.
XX
XX (UNOR-) UNIV OREGON HEALTH SCI
XX
XX Wong SM, Axthelm MK, Seattles RP;
XX
XX WPI: 2000-376552/32.
XX
XX New rhesus rhadino virus for producing non-human primate model useful
XX for testing potential treatments and efficacy of the candidate vaccine
XX for conditions associated with RRV infection -
XX
XX Claim 5; Page 123; 14pp; English.
XX
XX The present invention describes a novel rhesus macaque rhadinovirus
XX called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
XX RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
XX encoded by the genome sequence. The present invention also specifically
XX claims the individual open reading frame (ORF) nucleotide sequences from
XX the genome which encode the individual proteins, but these sequences are
XX not given. A non-human animal infected with RRV can be used for testing
XX the efficacy of drug in the treatment of condition associated with
XX infection with RRV such as Kaposi's sarcoma, lymphoproliferative
XX disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
XX hypergammaglobulinemia or autoimmune haemolytic anaemia, by
XX administering the drug to a immuno-compromised non-human primate
XX preferentially rhesus macaque monkey obtained by as a result of infection
XX by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
XX non-human primate model for testing potential treatments for conditions
XX associated with RRV infection. It is also useful for testing the
XX efficacy of the candidate vaccine against RRV infection or conditions
XX associated with its infection by administering the vaccine to the
XX subject capable of infection with RRV, inoculating the subject with RRV
XX and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
XX to AAB53213 represent sequence used in the exemplification of the

```

```

CC present invention.
XX
XX
XX Sequence 1132 AA;
QY
DB
QY 1 MENTQKTVTVPTGPGVYVACPVEDLDLEISFLAARSTDSDALPLMNLIVERTTS 60
DB 11 LEDNGDS-RAPIGAGGVYVAYSQKDEPFAEASTLGNPSSGVEFSLPILGLIVEHEFPL 69
QY 61 SL-AVVSAGATTGLACAGITLKLTTHSHFVPSVVFHCKKAVLPSSAAPNITRACNAAR 119
DB 70 TVAAAYKKVDTTLA-----VKVTC--FIREVIVFINASLFPFVDTGTGELNECEAFAL 122
QY 120 FGFSSRC--GSP--VQCAVETGAEICTRGLSEPENTILYVTAJLKE-----AV 166
DB 123 FGTOTPIFGPPPSIMNPLE-----CPQLP-DKDEMPICGVVTEGPKELMKGLVPAV 175
QY 167 FMCNVEPLHWGSLDVIHINSHQVPIPIPLFVQLFMPVNPPLVPDPFNTNHSISGEFVYPT 245
DB 176 FVLVQ-----VLAGKQAKKVFPLYDEDLFAPKGRHM-----P 207
QY 227 PRYNTGLCHLILHCCVLAAPAAVALRKAVNTAVAGAAHLADENHGAVALPPDITVITYQS 286
DB 208 RFYHKQVSAVLVDSLEFSLISQALRLKLVTAIVIAIEKQFMJLHYKIAKI---VQAKQF-- 262
QY 287 SSSGTTAPGAPPNQVNSTKSPSGGFEPRLASTMAADTALAEVIFNTGYEERTPTI 346
DB 263 -----STLPKPTTDDSGHMIYDSVAELALSYGCF-----LECHQDA 300
QY 347 KE-----WEMPTGMEETLPLNALASVTARVAVITGAAVSPNSALVILEVELSMTGA 400
DB 301 CELLNVDWPIFDGCDSPPEARVNALEPWSAQVAVHVAQGLFAANSVLYLTKVQKQAPFG 360
QY 401 KDSGPGSPFNKFTYQKAGPHLAANVQDRODH-----VLSQSSTGSSNTSPDYIALICG 455
DB 361 K--GDVNVVNSFLQHLQGLNLEATIKENSGEAFKQVPSNALDSSFTPY--HLAVNAS 415
QY 456 FGAPLLARLFLFLEPRDASAFGTGSHSDALKYVVTGTFSELPSCGEHTRPVCNHTVHR 515
DB 416 FSPHLLAKLCYVWQFLQHHSSSTNGAFNMVHYGTANSEKCTICHNTATCINTLPYK 475
QY 516 LRQMPREFGATRG-LIVGFTMNSQYSDCDPLGNAFYLILKPSDGT-EAAKATMDGT 573
DB 476 LKDRFPAVTTQRRDDPVYVGTAGT--FNDEILIGNFASFRDREDDGNADHEPKYTYWL 534
QY 574 YRATLEELFDLQERLLDAGACSSSGLS-----SVIVDHTPRRIIDTLARITQRT 626
DB 535 CQTVTEKL-----SALGITEDHDHNVNLTINQSLFVKRGIDSLVQGE 578
QY 627 TTQFMKVLVETRDYKIREGLSEATHSMALTFDPYSGAFPTTNFLVPTTHLAAVQDALS 686
DB 579 VMFVNSMIK-NMNPFRHKSVAHHILOPCCNVWQAPCAVILNLYKSLMI IQDILCP 637
QY 687 QTRVVRVYQ-----QVEGRNFRN-----QGVYLPREVDLF 718
DB 638 --YCMITEDNPMAGILPSEMLKMHFDTLWTFNKAAL-LDRKVLTVGTGLKTVHPRMFDFF 695
QY 719 -----NCGFISTGSIIVTISEGVSAPNPTIGQAPAGKIFLQGLAVASVAVIIDIWKN 773
DB 696 DTNAGSKGLMAPKMYK-----IAAMNVVERSIKIKN 724
QY 774 RVVFSNGTILSEAAARLV---GLASAVYQKQKPVMDLHAGLFLKQPHGLLEPKSMP 840
DB 730 KILFSN--TAGSRAVVSQGFAPKPGIKRQY-----VVAAPYMKFLNSJHRAFL----- 774
QY 831 ENSKSPRQWFMVLLQBNQMPADK-LTHEETITIAA-VKAFTEYALININLPPTGIGE 886
DB 775 PUKRTAALVLMHKISQNNKRPVLRKDVDDDELAELVSVYKNSLAFEEETNLDVVPDLSMS 834
QY 889 LAQFYMANLILKQCDHSQYLINILISITGAKRKPDSSVLIHWIKKQVISAADLEIQAKA 948

```


CC class switching. The isolated nucleic acid molecules and the peptides
CC they encode are useful as immunosuppressants, anti-viral agents or
CC vectors for gene therapy. These are particularly useful in protecting
CC humans from a variety of diseases, e.g. allergies, autoimmune diseases,
CC cancers, cardiovascular diseases, graft rejection, hematopoietic
CC disorders, immunodeficiency diseases, immunoproliferative diseases,
CC immunosuppressive disorders, infectious diseases, inflammatory diseases,
CC jaundice, septic shock, or other immunological defects, as well as other
CC genetic or metabolic defects.

SO Sequence 35 AA:

Query Match 2.9%; Score 183; DB 22; Length 35;
Best Local Similarity 100.0%; Pred. No. 3, 1e-10;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 545 DPLGNVAPYLIIKRCQDTEAAKATMODTYRATLE 579

DB 1 DPLGNVAPYLIIKRCQDTEAAKATMODTYRATLE 35

RESULT 8

AAB07562

ID AAB07562 standard: Protein; 1841 AA.

AC AAB07562:

DT 20-OCT-2000 (first entry)

DE Protein encoded by the bleomycin (BLM) gene cluster ORF30.

XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
XX bleomycin, bleomycin analogue, holo carrier protein, thiazolidine;
XX thiazoline, bithiazoline, microbial metabolite, sugar.

OS Streptomyces verticillius.

XX MO200040704-A1.

PD 13-JUL-2000.

PE 06-JAN-2000; 2000MO-0500445.

PR 06-JAN-1999; 9905-0115435.

PR 05-SEP-1999; 9905-0116848.

PR 05-JAN-2000; 2000US-0477962.

XX (BECG) UNIV CALIFORNIA.

PI Shen B, Du L, Sanchez C, Chen M, Edwards DJ.

DR WP1: 2000-465974/40.

DR N-PSDB: AAA58471.

PT New bleomycin gene cluster components useful for peptide and/or
PT polyketide metabolites, especially bleomycin, production and for
PT chemically modifying biological molecules.

PS Disclosure; Page 97-137; 162pp; English.

CC AAB07556-78 represent proteins encoded by open reading frames (ORFs)

CC 8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded

CC by the gene cluster are useful for producing peptides and/or polyketide

CC metabolites, especially bleomycin or bleomycin analogues. They are

CC also useful for chemically modifying biological molecules to produce

CC branched methyl groups, and for coupling amino acids and fatty

CC acids. They may be reacted with an apo-carrier protein and coenzyme A

CC to produce a holo-carrier protein. The BLM gene cluster or catalytic

CC domains can be used individually or collectively to produce

CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing

CC microbial metabolites. The BLM gene cluster may also be used to produce

CC sugars.

SQ Sequence 1841 AA;

Query Match 2.5%; Score 156.5; DB 21; Length 1841;
Best Local Similarity 19.6%; Pred. No. 3, 0.0013;

Matches 268; Conservative 133; Mismatches 470; Indels 493; Gaps 67;

OY 14 PLGVYACREDDLEISFLARST-DSDLALPLMRNLTVEKTFSSLAIV-----65

DB 461 IMLVVSARKIKRLALAAVKKLAAWSAPKEVTKLADLAATLAKGRAPVRAAVVCHDIPRA 520

OY 66 ----SCAR-TTGLAGAGITLKTTHSHYPSVYVHGCKHVLPSAAPHNLIRAGNAANERF 120

DB 521 ARLLGAGKGTALPGF-----EAVFLF-PGUGTLPPEDTGKGLVADVAFRAHF 567

OY 121 GFSRCGQPPVDCAVE---TTGAEICTRLGLEPENIIIVAVLAFKRAVMCVPLAHYCG 177

DB 568 -----DACACFPALGIDLHAALGPADDT--RAAPALFAVEVALAKTLMKMGV 615

OY 178 LDIIVINH--GDVIRIPLFVQLEMDVNLVDPENTHRSIGEGFVPTPEYNTGLCH 235

DB 616 RPAAMLGHSIGEVVAATLAGV-LSPDALILVBARAAQH-TMPGKMLAVPLPPDLRP 673

OY 236 LIHD-----CV-----IAPMAVALRVNRVTVAAAGAAHLADENHCAV-- 274

DB 674 LRPPEVEFSAFNAPGRGVGPEPEVAELRRLARRGVPAEELATAH-AF---HSAVER 729

OY 275 -----LPPDITVTFQSSSSG-----TTT-----ARGARR-- 299

DB 730 LIDGFRGVLGGVRLRPRLRYV---SLTGDMDAAVTTAYWLAHLRRVYRAADGIRK 786

OY 300 NDVN-----STKSPSSGGEERLAS-----IMAADTA---LHA-EVIFNTGIV- 339

DB 787 LDGVALVLEIGKAGITGLAKRAAGGEPVYIVKCLAAVDFAASLPHAAVALMRSCAV 846

OY 340 -----ETPTDIDK-----EMPTIGME 356

DB 847 DMTAFHRCGRPRRTVPGYPEQVRHWMIDAPDESEPTDLATLAKELRTGSDPLAVDQR 906

OY 357 -GTLPRINALGTYARVAVIGAMVSPNSALYL-TEVEFGMTEAKDGGPSFNFFQ 414

DB 907 PGLRTGLNRLCALARD-----YATGVEASGV-----POFHAFID 943

OY 415 FAGPHLAANDQDRDGHVSSSGTSSNTEFSVDYALIGFGAPLAPLLEFLERDAG 474

DB 944 YLRTLAASAPAAD-DAGTTIAETIAAH-----PSRSLGVDLLRHCAQG 985

OY 475 -----AFTGGHGDALKYVTGPFDEIPECSICEKHTRPCATTHRLRQHM 520

DB 986 YPALSTPGALDVLTPAGSGDLRRLTG-----ETADIRATGRL 1026

OY 521 PRPG-----QATROPIGVGTMSQYSDCDPLGNYPYLIIKRCQDTEAAKATMODT 573

DB 1027 TRLAGSLDLRLADRRGRPLRVLEAGAG---AGSLTQALVTRAPG-----RLDY 1073

OY 574 YRATLERLFI-----DLEQERLDRKAPCSSEGLS-----SYVDHPT 611

DB 1074 HAMDISHHFYALAGRAAKRGDLDFVRAKYLDIAKDQCEGDFAGKEPVLVCGILVVAHTPD 1133

OY 612 FRRILDTLRARIPOTTTOFVKLVETRDYKIRGLSATHTSMALTPPYGACAPIT--- 668

DB 1134 LKRTTLGHRSLMAPDGT---LALLET-----TADDPWLTMIGLTDGW 1173

OY 669 -NFLVAKTH-----LAVVODLALSQCHGVFGQGVFGNPNNO-----PQPLV 710

DB 1174 WHHTDRTRTHGLDLAAGWMLLAGEDPAIVADVIVPPGCPDAAIILARQPPRAAAPS 1233

OY 711 PRPFVLLF-----NQGFI-----STESITYTISGPPSAPAPPTLGGD 747

DB 1234 CKRDVGTWCYARQWRHAAAPADPATTCGCLLDGDDAKAVASRIE--ALGVIVTTVCGC 1291

OY 748 APAG-----RIFFDIDLARVSEV--IRDIVKNNVVSNGNCTINSEARARILVAGSAYOR 801

DB 1292 RPPGPEKRELVGPATRLAVDLMPLRDASHKRGAAAGAGVRTQDDALJHLHLHARAFGA 1351

UY 802 QERR... VDMHGLGFLKQFHLIFPRGMPNKSPPQWFWTLGQNPADKLT 856
 DB 1452 LERKHARAVTVTGA-----HNYL-----GDDLA 1376
 UY 857 DEETITL AAVKRETEY AAININLPETIGFAQFYMANILIKYTHSYLINTLT 913
 DB 1477 DEEHATVPAAAKVAPREYPMINCTALIVERGLDAE---RLADLIYRELGAAR---ETTV 1429
 UY 914 STETARR... PROSSVI HMLPKOVTSADILETGAKALLEETENIPHEMTAFSTHL 969
 DB 1440 TACRGRRRTTTP---VKQPLPA---PERPAVRPGGVYL 1463
 UY 979 VVAAANQRIYVAGISTEYFEVA GNNPVTGA INMSSTLWCKKAVFTLEFDP 1022
 DB 1464 WCGAGG...GLHAAVLCNARTTVVTHRRHPDAGCMGLDAGH---PAAVVR 1515
 UY 1024 RRTTACPRGRTGVTGSSNPPETLSTQ-----VPGITVSGG--- 1062
 DB 1516 LRSIAA...TGATVVVRADICTDHDMALADEVEGAGHPVRCVHAAGVPT 1565
 UY 1063 AMVQALATVAVRACARAOHMAFDIMLSLIDDEPLARDLEL 1105
 DB 1566 ASMIQRRKAGTUALAKLTGL-----VLDEVFAHRDLDL 1603

RESULT 9

AAH75958
 ID AAH75958 standard; Protein; 3210 AA.

AC AAH75958;

XX 29 MAY 2001 (first entry)

DE Mycelia sterilia cyclic depsipeptide synthase.

XX Mycelia sterilia: cyclic depsipeptide synthase; antihelminthic;

XX PF1022;

XX Mycelia sterilia.

XX W0200118179-AL.

XX 15 MAR 2001.

XX 07 SEP 2000; 2000MO JP06104.

XX 07 SEP 1999; 94JP-0254040.

XX 06 APR 2000; 2000AP 0164291.

XX (MEL) MELJI SHIKA KAISHA LTD.

XX Mideb N, Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T,

XX Althara S, Furumura T, Kleinkauf H, Murakami T,

XX WPL 2001:265979/27.

XX N P50B; AAT79702.

XX Novel cyclic depsipeptide synthase and gene encoding it for efficient

XX production of antihelminthic substance PF1022

XX claim 1; Page 59-80; 94PP; Japanese.

XX The present sequence is a cyclic depsipeptide synthase from Mycelia

XX sterilia. The present protein, or a protein containing an addition,

XX deletion and/or substitution of one or more amino acid residues is useful

XX for the efficient production of the antihelminthic PF1022 (cyclo(D-lactyl-

XX L-N-methylleucyl)-D-4-phenyllactyl-L-N-methylleucyl-D-lactyl-L-N-

XX methylleucyl)-D-4-phenyllactyl-L-N-methylleucyl)).

XX Sequence 3210 AA;

XX Query Match 2.0%; Score 121.5; DB 27; Length 3210;

Best Local Similarity 18.7%; Pred. No. 0.56;
 Matches 11; Conservative 79; Mismatches 227; Indels 177; Gaps 24.

UY 298 RNDVNSTKSPSGGHERBLASIMADYALHAAVFNFTGIYERTPPDIKRWMPICME- 356
 DB 2550 KRAVIAKRSQAAVA-----KPKVVDPEVNEAL-----CEFAEVLGMEV 2591
 UY 357 GTLPRIINALSYTARVAVGAVGAVFSPNSALYLTEVEDSGMTAKDGPPPSNRFQFA 416
 DB 2592 CVNDHPVQLGCHSLATKLVARLSRLNCRVSVHVDIQ 2630
 UY 417 GPHLANPOTDRDGHVLS-----OSTGSSNTEFSVDYALIGGFAPLLARLL 465
 DB 2631 PVISCLAVILPGQITLIFNAILFVIESGWEQTNSAFETISLDEAVLKEFAVLG 2686
 UY 466 FYLKRACAGAPFGCHG-----DALKYVIGTDSFPGSLCKKHTRPVCAHT 511
 DB 2687 VEVSTDSFEDLGHSIMATKLAAPSRLDVPYSIDIDHSVFLMAK 2737
 UY 512 TVHRLKRMRFEGATROPUGVEFTMNSGYSDCPDIGNAPYLL-----PKPSD 561
 DB 2738 ----IRLQAKGHEATN---GV-----QIANDAPQOLISVNDPELVORLIAP 2778
 UY 562 QTEAAKATMODTYRAT-LERLF-----IDLEGERLLDRGAPSS-EG 602
 DB 2779 QLCSPETILDVYPATQMGVHFLNPYTGKPSPTPHIDEPPD-----ADCAILMRAC 2832
 UY 603 SSYIVDHPFRRLIDPIRAFI EOTTQPMKVIVETRYKIRKGISVATHS----- 652
 DB 2833 ASLAKHFDITRYVLEAKGELYOVVLRKNDVPIEM--LQTEENINSTRSLDVADEKPT 2890
 UY 653 -----MALTFDPSGACPTTNFLVKRTHLAVVDLALS-----QCHGVFGQV-EG 699
 DB 2891 RLGPILRIALILEKPGS-----TLVYLRSLHALVGLSLEHILHSLHILFPGGSLPP 2944
 UY 700 RNFNRQFQPL--KRRVLDLPNGSF--STRSITVTLSEGPSAARPTLGAAPAGRTTSD 757
 DB 2945 PKFAGYMOHVAASSRRESDYWR--SVLRDSSMTYIKGNNTIPPEPPQUSTPSAHMAK 3003
 UY 758 IARVSEVIDIRKRVNVPSCNCT-----NISFARARIV---GLASAO 800
 DB 3004 VVITPQANTDSRITRATITTTACALMLAKEDNSDVYFGRTYSGRGLEPLAHQ 3057

RESULT 10

AAV00206
 ID AAV00206 standard; Protein; 1074 AA.

AC AAV00206;

XX 20-APR-1999 (first entry)

DE Enterococcus faecalis protein IP102.

XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

XX detection; attenuation; antigenic.

XX Enterococcus faecalis.

XX W09850554-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08959.

XX 14-NOV-1997; 97US-0066009.

XX 06-MAY-1997; 97US-0044031.

XX 16-MAY 1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Bailey C, Choi GH, Bromocky A, Kunsch CA;

WP1: 1999-070095/06.
N-PSDB: AAX20196.

New isolated *Enterococcus faecalis* polynucleotides - used to develop products for the detection of *Enterococcus* and for use in vaccines for prevention or attenuation of *Enterococcus* infection

Claim 9; Page 197-198; 301pp; English.

The present sequence represents a protein isolated from *Enterococcus faecalis*. The present invention describes genes, proteins and antigenic polypeptides isolated from *E. faecalis*. The proteins can be used in vaccines for preventing or attenuating an infection caused by a member of the *Enterococcus* genus in an animal. They can also be used for detecting *Enterococcus* antibodies in a sample. The nucleotide sequences can be used for detecting *Enterococcus* nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of *E. faecalis* protein activity.

Sequence 1074 AA:

Query Match 2.0% Score 123 DB 20 Length 1074:
Best local Similarity 18.3% Pred. No. 0.092; Mismatches 415, Indels 352, Gaps 52;
Matches 206, Conservative 150.

DB 67 KITSLKAEYNGAKUTVFC-IERGVSLPEVTHGY-----QKNL 104

UY 250 FVENYAVARGAHLAFCEHNEGAVLPOLITYFEQSSSSGTTIAGAPRNINSTSKPS 309

QY 310 PSCGFERRIASIM--AADTALAEVIFNTGIYE-----TPTDIKEMPMFI 353

DB 105 PSMSEKAKLVSYLWKEKAGTIDITNMVAQKIMEVNGYKLSIKRLGASGVIK----- 158

QY 354 GMEGFLPRLNALGSTAR-----VAVIG-AMVESPNALYTEVEDSGMTEAKDGP 405

DB 159 SIECKINK--ALEEYOKKSHFNITVKIILGOSTLIDNELNLSEFDVYVQNTA----- 211

QY 406 GEFNRFYQFAGPHLAANFQIDRLDGHVLSVSIGSSNTEFSVDYALIGCFGAPLLARLL 465

DB 212 ---NIDRVIGNQVLTP-----NSNSKSGTLTKKSGATGTP--VAYKK 251

QY 466 FYLEKCDGAFIQCGRDALKVYITGFDSEIPCSLCEKHTRPCAHTVYRLQRMRFQO 525

DB 252 AGLOTVMAGALDKPNTYAKINVEIKGS--LKIKKIDKESGDIVPETVFIH-----DFGR 304

QY 526 A-----TPQPIGVFGTMSQYSPDPLGNVA-----PYILIRKPDQTEAAKATM 570

DB 305 ALPSKDVTTDKGCI-----SLIDGIPGCTKVTITEKSVDPDPMIOTITMAATIKAGETI 358

UY 571 QDTYATLLEFLFIDLEGEKLLRGAFCSSSEGLS-----SVYDHP--FPRIIDTLRA 621

DB 359 SMTSKMRQKQILLEKTO-VEGTDLMDNNYSLAGNTPAIRKDSPAGEIYGEITDEK 417

QY 622 RLEQTTQMKVIVETRDVKIRGGLSEATHSMALTFDP-----YGCACPIITNPIVKKPT 675

DB 418 KAE--TPKELANALELGTYYTE--TKSSNGFVNITKPKIVELKYANQYALVATSNVKGQ 473

QY 676 HFAVVDIAL-----SQPH-VFYQYQV-----EIPNFRNQFVPLRPFVDL 717

DB 474 NGEITGEFTLTKEDKDTGNHSGKAEFKCAEYTLTAKDQAVKMSBAFK-----TEL 526

QY 718 FNGVFIETSTSTIVTLE-QPVSAFNPILVQ-----TAFAVIFTFDGLAVSVE----- 744

DB 527 VKGTASDETVTLLEKQVAVKHLAINVEYMOETKAPEGYTLDETKEYFVSILKKVUNNE 586

QY 765 ---VIRDIKVRKNRV-----FSGNCINLSEAA-----KARLVGLASAYQOEKR 805

DB 587 KNAVITRDVTAKQVIRFGEDEFKFAAGSAGTAEFGNDLSFKVSPLEGXETTGAEKDA 646

QY 806 VDMILGALCF-----LKQPHGLLEPPRGM-----PPNSKS-----PNP 838

DB 647 TTACNQLGFDGFGKFPENIPYGDYLLAEIRFA--PFCFOKITPLELIRSTPEKKNKDIYAKS 703

QY 839 QMFMTLLQNRQ-----MPADKLTHFEITTTAAVKRFEVEYALINFINLP--PTICIG 888

DB 704 EYVFITTEEGQKPIKMYVPEKLTNNE-----FVSINRLMLYDLEKESLSL 754

QY 889 LAQFYMANLILYCDHSHOYLNTLSITIGARRPRDSSVLHWIRKD--VTSADILENO 945

DB 755 LAIWKQGNKLTMLDPTB-LVYDKL-----RYNLHRIKEDMYVVAQALIVFA 799

QY 946 AKALLEKTENLEIMTATST-----HLVFAAMQPPMYVLISISIKYFA 992

DB 800 TTAADKEDKRAKPVYLAETLAILANKKIGYKILHKLAEQ-----VLDSIVLNFV 853

QY 993 ACNNRPFQAGNNSGNGCKNVCTLPFDRIKPIIACIQCQFICVYCHSSCKRETTLS 1052

DB 854 YENKAFEAGNE-----PVA-----KQASLNN 875

QY 1053 QVKGILVSGAMVQALIVAVHVAQAK-----AQHKAFLDMLSLIDDFLAKU-- 1102

DB 876 QAV-----TVNCTIERHWSIOTKALHEDQSOTFHQVMDMDVSVTHQVLD 923

QY 1103 --ELHDIILULETEFWTEGALEAVKILDE-----KTACD-----GETPINALFNLD 1149

DB 924 GSKKAEFTILYALLPUGTKKEIKMSKSKIEHVMKAEFTIKIVLAKVQICKYPPGCTIF 982

QY 1150 SCEPSHDTTSNV-----LNISSNISGTYVGLKRPEDDE 1185

DB 983 -TEINYEKGVNNGKHNEDEKRSQUTLPKEVPTIESTPEKPE 1024

RESULT 11

AAAY00188

ID AAAY00188 standard; protein; 1074 AA.

XX

AC AAAY00188;

XX

DT 20-APR-1999 (first entry)

XX

DE Enterococcus faecalis protein EF094.

XX

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

KW detection; attenuation; antigenic.

XX

OS Enterococcus faecalis.

XX

PN W09850554-A2.

XX

PD 12-NOV-1998.

XX

PF 04-MAY-1998; 98W0-US08959.

XX

PR 14-NOV-1997; 97US-0066009.

XX

PR 06-MAY-1997; 97US-0044031.

XX

PR 16-MAY-1997; 97US-0046655.

XX

PA (HUMAN-) HUMAN GENOME SCI INC.

XX

PI Bailey C, Choi GH, Bromocky J A, Kunsch CA;

XX

DR WP1: 1999-070095/06.

XX

DR N-PSDB: AAX20178.

XX

PT New isolated *Enterococcus faecalis* polynucleotides - used to develop products for the detection of *Enterococcus* and for use in vaccines for prevention or attenuation of *Enterococcus* infection

XX

PS Claim 9; Page 186-187; 301pp; English.

XX

CC The present sequence represents a protein isolated from *Enterococcus faecalis*. The present invention describes genes, proteins and antigenic polypeptides isolated from *E. faecalis*. The proteins can be used in vaccines for preventing or attenuating an infection caused by a member of the *Enterococcus* genus in an animal. They can also be

used for detecting *Enterococcus* antibodies in a sample. The nucleotide sequences can be used for detecting *Enterococcus* nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of *E. faecalis* protein activity.

XX Sequence 1074 AA:

Query Match 2.0% Score 123; DB 20; Length 1074;

Best Local Similarity 14.4%: prot No. 0.092; Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```

250 RYRNTAVARGAHLAFDENHGAVALPDITTYFGSSSGTTTAAARPNVNSTKPS 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 RTSLYAEVNGAKUTVEG-IEGVSIPTEVTHGY-----QKNPL 104
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 ISGGERKIASIM AADALAAEVFNIGTYEE-----TPPDIKEMPMI 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 RSMSSKAKLVSVLWEKAGTIDITNVAOKRMIEVNGYKLHSIKRLGASVDIK----- 158
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 SMGCHLPRIKALGVYAR -----VAGVIG-AMVSPNSALYIEVEDSGMTEAKIGCP 405
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 STECKINK ALIEYQKKSHPNHTIKTAGSTLLDKNELNLSHPDQVYQVNA----- 211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 GSPNKFYGFAPPHLAANVQTDKSHVLSSTGSSNTERSVDYALIGFGAPLLARLL 465
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 NIDYVIGNOVILITP-----NSNSKSGTLLTKKSAGCTTP-VAYKK 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466 FYLEPDCAFTYGHDAALKYVIGFTPSLPSICKEHPVCAHTYHRLHQRMPQO 525
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ANIGYMAVALDKPNTYALIKINVEIKGS-LKIKIDKESGDIPEVPHL-----DFGK 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
526 A-----TRQPIGVGCTNMSQVSDIPIQNYA-----PYLILKKPDIDTAAKAIM 570
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 ALPSKIVITDKCI-----STLHGIPHCQKVIIEKSVDPYMIIDITPMALTIKAGETI 358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
571 QUTYATLTERLEFDLPOELDKGACGSSGGS-----SVLDIHT-----FRILDTLTA 621
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 SMTSKMMKQGLILKQ-VEGTGIMMDNYSIACNTFAIKDSAGELVCHITTDKGO 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
622 KREOTITQMKVIVELIKYKIKRQISALHSMALIFDP-----VSGACPIITNPLVKRI 675
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 IAF-IPKELANALDELGYYTE-TKSSNGEVNTEKPTKVELKYNQVALVTSNVKRG 473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
676 HAAVVDIAL-----SOCHVYFYQGV-----FCHNFPNGQPIVILKRFVIL 717
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 NQETIGETTLIKEDKQNGESQKAFKGAFTLTAKDQAVKMSAPK-----TEL 526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 VKGTASDDELVALDEKQVAVKHALINFYMQETKAPQVITDDETKYPSIKKQDNNE 586
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 VIRDIRVNNRV-----FSGNCTNLSSEA-----KALVGLASAYQNRKR 805
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
587 KNAVITRDTAAEYIKRGEDEFKAGSADGTAETGFENDLSKVSLEGTIXETIAGEDA 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
806 VIMELGALCP-----ILKQFHGILFPGM-----PPNSKS-----PNP 838
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
647 IACNPGUQCHIKRYKFNENFYQGYLLELHA-----FPGHCKIIPLEIKRSIKKKQDYAAS 703
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 QWEHTILQNRQ-----MPADKLTHERITTLAAVKRTEEYAAINFLNP--PTGIGE 888
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
704 EYVFTIIEGQVQPIINMVAIPEKILNNP-----FVSINIKMLIYDDEKRLDLS 754
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
889 LAQFPMANILIKYCHRSOYLINTSITGAKRPPRSSVLMIRKD---VISAAILEQ 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 LATMKRNNKILNTLQTE-LYDQL-----RYNLHLEIKEDWYVQAQIDVEA- 799
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
946 AKALLEKTENIPELMIAFIST-----HIVPAANQRMVIVIGISISKYHQA 992
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
800 YKAAGPKPKKAPVYIAETATLANKKRTGKLLKILAKQ-----VLDKSTVLPNV 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
994 ANNNVAFQAGNMSLNCKRNV-PLFTFDRTRFIIACPRKGFICHPVTPSSGNRETTLS 1052

```

```

DB 854 YEKRVAFENAE-----PVA-----KIASLN 875
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1053 YVQGLIVSGAMVQALVAVRAVAK-----AQHAFIDWISLIDEPILAKIL- 1102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
876 QAO-----TVNCTIERHVSLOTKAHLEDGSQPTFHIDVMDPDDVSVTHVLD 923
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1103 ---EELHDIITQLETPEWYVGALEAVKILDE-----KITAD-----GETPINLAFNF 1149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 GSKFAFTTILYALLDGTNKPTRKSKTTHENVNKEFTVIAKRYNGKVPYDIAEIR- 982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 SCEPSHDTTSNV-----LNISSNIGSTVPGLRKPEDE 1185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 983 -TEINVEKQGNVNGKIINEDLKEKSQTLTPKEVPTISTPQPE 1024
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12

ABP43407
ID ABP43407 standard; Protein; 1074 AA.

AC ABP43407;

XX 05-AUG 2002 (first entry)

DE E faecalis EF094 protein.

KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

OS Enterococcus faecalis.

PN US2002045737-A1.

PD 18-APR-2002.

PF 04-MAY-1998; 98US-0071035.

PR 04-MAY-1998; 98US-0071035.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;

DP WPI: 2002-425450/45.

DR N-PSDB; ABN98163.

XX New genes and polypeptides from *Enterococcus faecalis*, useful as

PT vaccines for preventing, treating or attenuating an infection caused by

PS a member of the *Enterococcus* genus in an animal, particularly *E.*

XX faecalis

XX Claim 9; page 163; 25pp; English.

XX The present invention provides the protein and coding sequences of a

XX number of polypeptides from *Enterococcus faecalis*. The proteins can be

XX used as vaccines for preventing or attenuating an infection caused by a

XX member of the *Enterococcus* genus in an animal, particularly *E. faecalis*.

XX The polynucleotide is also useful for preventing or treating *E. faecalis*

XX infection. The present sequence is a protein of the invention.

XX Sequence 1074 AA:

Query Match 2.0% Score 123; DB 23; Length 1074;
Best Local Similarity 18.3%: prot No. 0.092;
Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```

250 RYRNTAVARGAHLAFDENHGAVALPDITTYFGSSSGTTTAAARPNVNSTKPS 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 RTSLYAEVNGAKUTVEG-IEGVSIPTEVTHGY-----QKNPL 104
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 ISGGERKIASIM AADALAAEVFNIGTYEE-----TPPDIKEMPMI 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 RSMSSKAKLVSVLWEKAGTIDITNVAOKRMIEVNGYKLHSIKRLGASVDIK----- 158

```

```

QY 354 GMEGTLPRLNALGYSTAR-----VAGVIG-AMVFPNSALYTEVEDSGMTEAKDGP 405
DB 159 SIEGKLNK--AIEEYOKKPSFHNITVKTLLGGSTLLDKNELNSLSPDQVONTA----- 211
QY 406 GFSFNRPFYQAGPHLAANPQTDNRDCHVLSSQSTGSSNTEFSVDYALALCGPGAPLLAKLL 465
DB 212 ----NIDYRVIGNQLVLP-----NSNSKSGTLLTKKSAAGTGP-VAYKK 251
QY 466 FYLERGCGAGFTGGHGDALKYVIGTFDHLPGSLCKHTRKVCANTFVKRLQRMKRRKQ 525
DB 252 AGLQTVWAGALDKPNTYAIKINVEFKGS-LKIKKIDKESGDIIVETVPHL-----DPCK 304
QY 526 A-----TRQPIGVEGINSYSDDEPLGNVA-----PYLLLRKPDYTEAKATM 570
DB 305 ALPSKDVTTDKDGI-----SLIDGIPHGTQVTFPKSPDPYMTDTPMAATIKAGETI 358
QY 571 QNTYRATLPLPLTIDRDEPRILDKAPCSSSGIS-----SVIVHPT----FRRILITLTKA 621
DB 358 SMTSKNRKRGKQILLERTG-VEGTGLDMNDYSLA;NTFAIKRQSPASEIVAEITTEKQ 417
QY 621 KLEPUIIYQPMKVLVETIKQIKKEGISEATHSMALIFDP-----VSGACQPIINFLVKRI 675
DB 418 RAE--TIKRLANALHICITYVTE--TKSSNCYVNTFKPKVEIKRANQIVALTSNVKGQ 473
QY 676 ILAVVQDUAL-----SQCICVEYGVV-----EGKNEFNQHPVLRRRFVDL 717
DB 474 NOEITIGETLLTEKDKDQCNHSGKAEFKCAEYLLFTAKDQAVKWEAK-----LIL 526
QY 718 FNGGFTSTRSTIVTISF-GVVSAPNPTLQO-----DAPAGTRPDILARVSV----- 764
DB 527 VKGTASDETVLALDEKNQVAVKHILATNEYFMQETKAEQGTLDTEKYPVSKKYVNNK 586
QY 765 ----TIRIDRVKRNKV-----FSGNCTINLSEAA-----KALVGLASAVQKQEK 805
DB 587 KNAVITRDVTAKEQVIRPGDFPKFAGSADGTAETGNDLSKVPLEGTXEITGAEKDA 646
QY 806 VMLGALGF-----LLKQFHGLFFPRGM-----PPNSKS-----PNP 838
DB 647 TTACNEQQLGFDYQKRENI-PYGDYLLLEIEA--PEGFOKITPLEIRSTFKENKDYAAS 703
QY 839 OMFWTLILQHNQ-----MPADKLTREIEITTTAAVKRFEVEEYAAINFLNP--PFCIGR 888
DB 704 EVVFTITEGQOPKIMVTVPEYKELNNE-----FVSINRLMIDLPEKESLTS 754
QY 889 LAQFYMANILKYCDHSQYLINTLISITIGARRPDPSSVLMIRKO--VTSADILETO 945
DB 755 LATWKNGKRLMTIDPTE-LVTKI-----RYNLRIKEDWVVAQAATVFA- 799
QY 946 AKALLEKTENLEIMTATST-----HLVRAAMNQRMVYLGISISKYHGA 992
DB 800 TKAADKDKKAKIVVIAETATILANKKGTQWKILKRLTAEO-----VLDKSTVLPNV 853
QY 993 ACNNRPFQAGNNSGLNGCNVCHLFFPDKTRKFLACPKQGLCPVYTGSSNGRRTTISD 1052
DB 854 YENKVAPEAGNE-----PVA-----KDSALNN 875
QY 1053 QVRGILLVSGCANVQALVATVRAVAGAR-----AQHMARFDMLSLTDDERLARDL- 1102
DB 876 CAGC-----LVNLTIFRPPVSLQIKAHLEHSGQTFTHGIVMIFQIVSVHIVLV 923
QY 1103 --EELHDOITLOTLETPMTEGALEAVKILDE-----KTTAGD-----GETPINLAFNFD 1149
DB 924 GSKFAETITLIALILQIKKEIKMSCKLHREVMKEFKIKVLAKEVYICQNTFAGTKIF 962
QY 1150 SCPPSHDTTNSV-----LINSQSNISGSTVPLGLKRRPPDDE 1185
DB 983 -TEINYEKDNVNGKHNEIDLEKESQTLTPEVEVPTIESTPQPE 1024

```

PSOUT 13
 ABP43425
 ID ABP43425 standard; Protein; 1074 AA.
 XX

```

AC ABP43425;
DB 05-AUG-2002 (first entry)
XX
DE E faecalis EF102 protein.
XX
KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.
XX
US Enterococcus faecalis.
XX
TN US2002045737-A1.
XX
DB 18-APR-2002.
XX
PE 04-MAY-1998; 98US-0071035.
XX
XX 04-MAY-1998; 98US-0071035.
XX
XX (HUMAN-) HUMAN GENOME SCT INC.
XX
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
XX WPI: 2002-425450/45.
XX
XX N-PSDB: ABN98181.
XX
XX
XX New genes and polypeptides from Enterococcus faecalis, useful as
XX vaccines for preventing, treating or attenuating an infection caused by
XX a member of the Enterococcus genus in an animal, particularly E.
XX faecalis.
XX
XX Claim 9; Page 177; 255pp; English.
XX
XX
XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly E. faecalis.
XX The polynucleotide is also useful for preventing or treating E. faecalis
XX infection. The present sequence is a protein of the invention.
XX
XX
XX Sequence 1074 AA;
XX
XX Query Match 2.0%; Score 123; DR 23; Length 1074.
XX Best Local Similarity 18.3%; Pred. No. 0.092;
XX Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```


DB 575 ---SCDVGK-----TSHKRAMLVDOQINMFIPLNSMEGN----- 610

QY 1062 GAMVLAIVATVRAVGARAHMAD-----WLSLIDDEFLADELIDLIQFLRI 1115

DB 611 GAKLELAHSISEAVLGFEGMAIDNPNNLITWEAFSD--FAMAAIILITFIVSSET 668

QY 1116 PWTIVGAL-----EAVKILAEKFTACDGET 1140

DB 669 KMWESNALVMLPHIGYDGAASEHSSCRLERFQLCDSKETSADGUS 714

RESULT 15

AAV53665

AAV53665 standard; protein: 2387 AA.

AAV53665:

22-FEB-2000 (first entry)

Mechanical stress induced protein 608 amino acid sequence.

Mechanical stress; gene library; protein 608; osteoporosis; bone density; bone development.

Rattus sp.

Key Location/Qualifiers

Misc-difference 7 /note- "encoded by CAT"

Misc-difference 363 /note- "encoded by GGA"

Misc-difference 1453 /note- "encoded by NTT"

W09960164-A1.

25-NOV-1999.

14-MAY-1999; 99WO-0511066.

15-MAY-1998; 98US-0085673.

(OJAK-) OJAK BIOTECH INC.

Einat P, Mor O, Skallier R, Feinstein E, Fierman A:

WPI: 2000-054404/04

N-PSDB: AAZ36322.

Identification of stress induced genes for determining risk and preventing, treating or controlling osteoporosis

Claim 32, Fig 5A-Z, 308pp, English.

The present sequence represents protein 608, which was identified using the method of the invention, after subjecting rat osteoblasts to mechanical stress. Expression of the 608 gene was found to be upregulated by about 3-fold in cells subjected to mechanical strain. The specification describes a method for the identification of genes responsive to a specific mechanical stress. The method comprises applying the mechanical stress to an organism (tissue or cells comprising bone cells), isolating the specific cellular fractions and extracting RNA from them, and differentially analysing the RNA in comparison with control samples. The method is used to identify genes whose expression is responsive to a specific stress. The identified genes are employed in determining risk associated with a physiological or disease state. The risk determination methods are used for testing a medicament for gene therapy, these medicaments, or genes identified by the method of the invention, are used for treating, preventing or controlling a physiological or disease state (especially osteoporosis or bone density or other factors causing or contributing to osteoporosis or its symptoms or other conditions involved in mechanical stress or its lack. The methods can also be used for advancing research or studies in bone

CC development.

XX Sequence 2387 AA;

SV Query Match 1.9%; Score 122.5; DB 21; Length 2387, Best Local Similarity 18.6%; Pred. No. 0.42; Matches 145; Conservative 114; Mismatches 286; Indels 233; Gaps 40;

QY 416 AGPHLAANPQTD---RDGVLSQSSTGSSN-----TEFSVDYIALIIG 455

DB 64 SGAFLOCTKPTIPDSUKSKSLVTOEDNGSASTSPQDIEPFGSISLMTLSGNKADWV 123

QY 456 PCAPLLARILFVLEKDCAGFTGHDALKYVGFPSDEIPCSLCCKHTRPV----CAH 511

DB 124 IQRP-----SRSPAFTEENDYIM--LNASTNLVGVVDVNHQOPWQLLAIS 172

QY 512 TVHRLRQMP-----KEGQATROPVIGVGTNNSVSDCPDLAGNAPYILKPGD 561

DB 173 DSEPLLEKRPOLTEIPSSSKYKQVALRPEDLFTSLIAVVK-ADPWPQCKIVQINRT 231

QY 562 QTEAAKATMODYRATLEKFLIDLEGERL-----LIR-----GAPSSSEJ 601

DB 232 ATTLSTLIQVSTDAIALPRAEMPAEFLKWTMI--MMNNKLEFVIVASTIALSPK 291

QY 602 ---LSSVIVDHPFR-----RILDTLAKILF-QTTOPMKVL---VETRD----- 639

DB 292 DSPHLEMLLACSKVPAFVYSEFPIIDCKNKLLELMASFFAHLHCISTEDADAV 351

QY 640 ---KIR--EILSEATSMALTFDPYSAGF---CPTNFLVKRTHLAVYDGLALSGCHV 691

DB 352 LTVRTIVVEYCESTHDSGOHTVGTETLDPLCLSTG-----VIDASIS---WI 398

QY 692 FYGOQVEGRNFMQFVLRREFDENGEISTRTIVTLSEGPS--APNP----- 743

DB 399 LPGNTEFQSPDRD-----QIINOTLRILOV--PKDGHVQCVANSGADESS 447

QY 744 -----LGDPAPRGTDFDGLARYSVEVIDIRKKNVAVSGNCTINSEAKAR 791

DB 448 FKVSVOKKGQRMWEHREAGSGSLGE--PNSSVSLCKOPASLK-----LSASALTSGEAKO- 501

QY 792 LVGLASAVYQCKEKVLMNLGALG--FLIKVFGILFEKMPNSKSPNPWFMTLQ---R 847

DB 502 ---VSGVHRKKNHRLIHRRGDSTLRPRR--HRRQPLISARIPDQWAAILEKAK 555

QY 848 NQMPADKLTHEITT-----IAVKKFTEEYALNFINLPFCISLQAFYMANILK 900

DB 556 NSVP-----KKQENTYKPVPLAVPELITDEKDSQMI--FP----- 592

QY 901 YCHSQYLNLNLISLTGAKRPKPPSSVLMIRKDVSAADIEQAKALEKTENLEPM 960

DB 593 ---DEEFVLTAKSGVGRKSPADSGVNNGFMTISAGIEVIVNPQTLQ--SEHLPDK 649

QY 961 TTAFTSTHLVRAAM-----NQRPMVVLGISISKYHAA--NNKVPQA--NWSG-- 1006

DB 650 LFSVTINQAVTKASGNPSIAKIEDTQNP111FP--SVAFIPDQA-----QADRASSQS 702

QY 1007 ---LNGKKNCPLEFTRTRPFIACPRGGFI-----CPVTG---PSSGN 1045

DB 703 AHPVAGN---MATYGHNTYISFTSKASTVLQPLNFTESYGVQVITTVGSRPSSD 756

Search completed: March 28, 2003, 13:35:35

Job time: 96 secs



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 28, 2003, 13:33:59 ; Search time 31 Seconds
(without alignments)
1141.799 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294
Sequence: 1 MENTOKTYVTGTPGLGVYA.....DELFDLSGIFIKHGNITMEM 1203

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 265574 seqs, 29422922 residues

Minimum number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3054	48.5	1452	4	US-09-127-227-2
2	183	2.9	35	2	US-08-807-332B-9
3	183	2.9	35	4	US-09-338-876-9
4	131	2.1	757	4	US-09-413-814-84
5	123	2.0	1074	4	US-09-071-035-358
6	111	1.8	5215	4	US-09-071-035-394
7	111	1.8	5215	4	US-09-071-035-394
8	106.5	1.7	3567	3	US-07-642-734C-4
9	106.5	1.7	3567	3	US-08-439-009A-4
10	106	1.7	774	1	US-07-731-157A-7
11	106	1.7	774	2	US-08-541-780-7
12	105	1.7	774	1	US-08-633-760-46
13	105	1.7	774	1	US-08-633-760-48
14	104.5	1.7	635	4	US-08-931-608A-5
15	104.5	1.7	774	1	US-08-019-870-5
16	103	1.6	774	1	US-08-019-870-3
17	103	1.6	774	1	US-08-633-760-44
18	102	1.6	729	1	US-08-070-165P-6
19	102	1.6	729	2	US-08-885-418-6
20	102	1.6	774	1	US-07-747-801A-3
21	102	1.6	774	1	US-07-945-312-3
22	102	1.6	774	1	US-08-633-760-50
23	101.5	1.6	773	1	US-08-019-870-1
24	101.5	1.6	773	1	US-08-019-870-6
25	101.5	1.6	774	1	US-08-019-870-8
26	101.5	1.6	774	1	US-08-019-870-11
27	101.5	1.6	774	1	US-08-633-760-52

28	101	1.6	1024	4	US-09-562-737-4R	Sequence 48, Appl
29	101	1.6	7257	3	US-09-335-408-5	Sequence 5, Appl
30	101	1.6	7257	4	US-09-568-102-5	Sequence 5, Appl
31	101	1.6	7257	4	US-09-567-969-5	Sequence 5, Appl
32	101	1.6	7257	4	US-09-568-480-5	Sequence 5, Appl
33	101	1.6	7257	4	US-09-568-486-5	Sequence 5, Appl
34	101	1.6	7257	4	US-09-568-472-5	Sequence 5, Appl
35	101	1.6	7257	4	US-09-567-899-5	Sequence 5, Appl
36	100.5	1.6	1321	5	US-08-261-822A-3	Sequence 3, Appl
37	100.5	1.6	1321	5	PCT-US95-07744A-3	Sequence 3, Appl
38	100.5	1.6	2616	6	US-08-328-322-5	Sequence 5, Appl
39	99	1.6	915	1	US-09-036-987A-3	Sequence 3, Appl
40	98.5	1.6	2152	4	US-09-370-700-3	Sequence 3, Appl
41	98.5	1.6	2152	4	US-08-451-822A-13	Sequence 13, Appl
42	98	1.6	821	2	US-08-323-430-13	Sequence 13, Appl
43	98	1.6	821	4	US-08-471-570-8	Sequence 8, Appl
44	97.5	1.5	769	1	US-09-562-737-50	Sequence 50, Appl
45	97.5	1.5	1024	4	US-09-562-737-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-09-127-227-2
Sequence 2, Application US/09127227
Patent No. 6399354
GENERAL INFORMATION:
APPLICANT: David M. Knipe
APPLICANT: Travis J. Taylor
APPLICANT: Elizabeth McNamee
TITLE OF INVENTION: Replication-Competent Virus Expressing A
FILE REFERENCE: H098-05
CURRENT APPLICATION NUMBER: US/09/127,227
CURRENT FILING DATE: 1998-07-31
NUMBER OF SEQ. ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1452
TYPE: PRT
ORGANISM: herpesvirus
US-09-127-227-2

Query Match 48.5% Score 3054; DB 4; Length 1452;
Best Local Similarity 49.8%; Pred. No. 1.8e-304;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

QY	1	MENTOKTYT---VPTGPGYVY--ACRYEDLDLEISFLARSTDSIALPLMKRLTVE	55
DB	1	MENTOKTYT---VPTGPGYVY--ACRYEDLDLEISFLARSTDSIALPLMKRLTVE	55
QY	56	KTFSSLAIVSCARTYTGACACITLKLTHSHFYSPVFPFGCKHVLPRSSAPNLTTPACNA	115
DB	56	KTFSSLAIVSCARTYTGACACITLKLTHSHFYSPVFPFGCKHVLPRSSAPNLTTPACNA	115
QY	59	SGFEFNVAIVVSSRTTGGTAVSLKLTPIHSHSSVVFHGGRIHDPSTQMANLTRCLER	118
DB	59	SGFEFNVAIVVSSRTTGGTAVSLKLTPIHSHSSVVFHGGRIHDPSTQMANLTRCLER	118
QY	116	ARERGFSCQCPVDDGAVETGAEICTRLGEPNTILYLVVTALEFAVPMCVPLAY	175
DB	116	ARERGFSCQCPVDDGAVETGAEICTRLGEPNTILYLVVTALEFAVPMCVPLAY	175
QY	119	ARRHGFISYTRPPDILKHETTGELGELGIDPPRALILYLVVTRFPEACVACINNTFHL	178
DB	119	ARRHGFISYTRPPDILKHETTGELGELGIDPPRALILYLVVTRFPEACVACINNTFHL	178
QY	176	GSIDLVTHNSGVIRLPLEFVOLFMEDVNRVLPDFNTHNSLGESEFYFFPYFTGTGLH	235
DB	176	GSIDLVTHNSGVIRLPLEFVOLFMEDVNRVLPDFNTHNSLGESEFYFFPYFTGTGLH	235
QY	179	CGSDKVTTCGACVHRIPVPIQLFMDPSVPAEPFNAHRSIGNFYPLDPEFNRPLNR	238
DB	179	CGSDKVTTCGACVHRIPVPIQLFMDPSVPAEPFNAHRSIGNFYPLDPEFNRPLNR	238
QY	236	LHDCVIAVMAVALPYRNVTAARGAHILAPFNNHGAVALPDIYTYFOSSSGTTFAR	295
DB	236	LHDCVIAVMAVALPYRNVTAARGAHILAPFNNHGAVALPDIYTYFOSSSGTTFAR	295
QY	239	LLEAVVGAALVALKRRNDVAAARAAHLAPDENEGALALPDITFAEVSUG--KLPK	296
DB	239	LLEAVVGAALVALKRRNDVAAARAAHLAPDENEGALALPDITFAEVSUG--KLPK	296
QY	296	GARRNDVNSTKPSDGGFERRPASTMAADTALAHAFVINTICITYETTPIDIKEMWICM	355
DB	296	GARRNDVNSTKPSDGGFERRPASTMAADTALAHAFVINTICITYETTPIDIKEMWICM	355
QY	297	CGR-----DCCGCKPGAGCFEORIASVMAGDAALALEISVSMVFPNPPPTDLSAMPICGQ	351
DB	297	CGR-----DCCGCKPGAGCFEORIASVMAGDAALALEISVSMVFPNPPPTDLSAMPICGQ	351
QY	356	EGTLPLNLGLSTAVAVAGVIGAMVSPNSALYLTEVEDSGMTLAKDGSGPSFRFVQF	415
DB	356	EGTLPLNLGLSTAVAVAGVIGAMVSPNSALYLTEVEDSGMTLAKDGSGPSFRFVQF	415

```

100 452 D1AAAAANAVGATLARAAGIYGAIVFSTNSALHLEVDIAGIADPKDHSK-PSYHFFIV 410
101 416 ASPIHLAANPOTDDEHVL-----SSQSTGSSNTEFSVDYALIGFAPPLARLLFYL 468
102 411 PCHHVAANQVIRKRVNVPKRPAPLVGGIYQ-PPACHHIAADGFSALLAKMLFYL 469
103 469 ERTDAGAPITGHC DALKVYTCIPISEIIPCSICEKTHRVCAHTTVH QIKOMRRCQAT 527
104 470 EFKDGVIVGRQEMQVFRVADNSQTDVPCNCTFTFRHACVHTTLMRLARHPKVASNA 529
105 528 KQPIGVGIMNSYSKSDIQLANAPYLIIKQIQVTEFAKATIMQDIYKATLEHLPIDHQ 587
106 540 KALICVPGIMNSYSKSDIQLANAPYLIIKQIQVTEFAKATIMQDIYKATLEHLPIDHQ 588
107 588 EMLIDGAPESSEISVIVIEHFEFRIIDLELRLARLEOTTQFMKLVETROKIREGSL 647
108 589 QGYVQAVITAMCHLEIITINHAIHIVANNVQVVDKRVQIMRNLYGSRNKRKFDIG 648
109 648 EATHSMALLFDYVSCAPPTINIVAKRTHIAVVDIALISQCHCVYQGVQVCEHFNPOQ 707
110 649 FANIMSLIDPYATGCPCLDILGRSNLAVYQDLALSQCHCVYQGVQVCEHFNPOQ 708
111 708 IPIIRHFFVLPFGCPSTSTVTITSEG-IVSAPNPTIGDAPACRPGDILARVSVVI 766
112 709 IPIIRHFFVLPFGCPSTSTVTITSEG-IVSAPNPTIGDAPACRPGDILARVSVVI 768
113 767 KOLIKKRVVESCNTNLSSEARARIVLQASAYQPEKRVDMHIALGFLKQFHLLEP 826
114 769 KILKAKSVILFANASANASFAKAVASIOSAYQKPKRVHIIIGLQFLKQFHAIFFP 828
115 827 KMPNRSKSNQVWFTLQORQVADKTHRETTIAVKKRTEYAAININLPPTCI 886
116 829 NKPIPSNQNQVWFTLQORQVADKTHRETTIAVKKRTEYAAININLPPTCI 888
117 887 GELAPYMANLIIKQVDSOYLINILSTIIQAKRPROSSVILHVRKQVISAADIEVA 946
118 889 SELAYMANNOIIRYGDHSFYINILTAIAGSRHPSVOAAAM---SAQCGAGIIFACA 945
119 947 KALLKTEMLPELITATSTHIVRAAMORPVVYLGISYHGAAGNNRVQVAGNWSG 1006
120 948 KALIMAVIAHAPAMISMAASCNILKRVMAKRVYIGISYKYGAGMDRVQVAGNWSG 1005
121 1007 INCKNVCTIPIFDEKRFELIACPGCGIPIVTCSSCKRFTTSNOVGCIIVSCGAMQ 1066
122 1006 LMGKNATFELLIFORTKRVYLVAPRAGVYVAAISNLGGAHSSLECEGLKGIISGGAAYA 1065
123 1067 LAIVAVRAVAGARAHMAIDMISITIDDEIARQLEELHQIIOITTPVIVESAL--- 1123
124 1066 NSVAVATKSHDRIGQIPIVWIALDEYISSEPMELIARALKNGCNSWSDALIEVA 1125
125 1124 EAVKIDDEKTVAGGPTPTNIAFNFO--SFPSPHTTSNVINISGNSISGTVPLKRP 1180
126 1125 HRAKALVSOIAGAG-----VFNFQDQGR-----DINATPGCGIAPGAPAGAKRKA 1173
127 1181 PEDDELFDISLPEKKNIMEM 1203
128 1174 FEGDDEFG EEPHKKGLTMM 1195

```

```

RESULT 2
US 08-807-332B-9
Sequence 9, Applied for US/08807-332B
Patent No. 5959074
GENERAL INFORMATION:
APPLICANT: Dreyfus, David H.
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross
STREET: 1700 Lincoln St., Suite 3500

```

```

? CITY: Denver
? STATE: CO
? COUNTRY: U.S.A.
? ZIP: 80203
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/807,332B
? FILING DATE: 28-FEB-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Kovarik, Joseph E.
? REGISTRATION NUMBER: 33,005
? REFERENCE/DOCKET NUMBER: 2879-39
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 303/863-9700
? TELEFAX: 303/863-0223
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-807-332B-9
?
? QUERY MATCH 2.9%; Score 183; DB 2; length 35;
? Best local Similarity 100.0%; Pred. No. 1,1e-11;
? Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 1 DPLGNVAPYLLIKRPGDQTEAAKATIMQDIYKATLE 35
US-09-338-876-9
Sequence 9, Application US/09338876
Patent No. 6187584
GENERAL INFORMATION:
APPLICANT: Dreyfus, David H.
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,876
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,332
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700

```

TELEFAX: 303/863-0223
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 35 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-338-876-9

Query Match 2.9%; Score 183; DB 4; Length 35;
 Best Local Similarity 100.0%; Pred. No. 11e-11;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 DPLSNVAPYLIRKPGDTEAKAIMODYRATLE 579
 1 DPLSNVAPYLIRKPGDTEAKAIMODYRATLE 35

SEQUENCE 4
 -09-413-814-84
 Sequence 84: Application US/09413814

GENERAL INFORMATION:
 APPLICANT: Gesellschaft fuer Biotechnologische Forschung MBH
 APPLICANT: Bristol-Myers Squibb, Co.
 APPLICANT: Beyer, Stefan
 APPLICANT: Bloecker, Helmut
 APPLICANT: Brandt, Petra
 APPLICANT: Cino, Paul M
 APPLICANT: Dougherty, Brian A
 APPLICANT: Goldberg, Steven L
 APPLICANT: Hoelle, Gerhard
 APPLICANT: Mueller, Joachim
 APPLICANT: Reichenbach, Hans
 TITLE OF INVENTION: heteropolypeptide compounds
 TITLE OF INVENTION: heteropolypeptide compounds
 FILE REFERENCE: PCT/US 99/23535
 CURRENT APPLICATION NUMBER: US/09/413,814
 CURRENT FILING DATE: 1999-10-07
 EARLIER APPLICATION NUMBER: DE 198 46 493.2
 EARLIER FILING DATE: 1998-10-09
 NUMBER OF SEQ ID NOS: 107
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 84
 LENGTH: 757
 TYPE: PRT
 ORGANISM: Sorangium cellulosum
 -09-413-814-84

Query Match 2.1%; Score 131; DB 4; Length 757;
 Best Local Similarity 21.1%; Pred. No. 0.00079;
 Matches 146; Conservative 84; Mismatches 251; Indels 212; Gaps 37.

DB 54 VKITFTSSIAVVSQAT-----IGLACGTLTKLTHSPVSVVHGCKHVLPSAPNI 109
 DB 129 LERTLPRAVSIADAAAEALTSVAVGIVAS-----LPASAAA---167
 DB 110 TRACNAAPFEPFESRQGPVDAVETTSACICTRLGLEPENTILYLVATLKEAVEM 169
 DB 168 -----AAIQRLKRLMADIGS--PGPIHGCA-----ALKRPSV-----198
 DB 170 NVELHY-----GGLDVIHINRDEVIRPLEVGLFMPVQKRIVPPFNTHRSIRGPPV 244
 DB 199 -AFLOYTSSTGEPKVMILHNL-----LHNSRLTAHGFDLTSPDPV-----GVTW 244
 DB 225 PVPVNTGIC-----HIHDCVIAQMAVANR-----VRNTAV-----AGSAHLAPD---267
 DB 245 LPPIYDMGLIGILQALYRIRIVALLSPISFQRPKRWRAVSAAGASVSGGDFEAYDLC 304
 DB 268 -----ENHIGAVLPDITTYFGSSSSGTTTARGARRDNVNSTSKSPSSGFEKRLASIMA 323
 DB 305 VKSSSEFFPAALD-----IKSWFVAFPGADPVRADTLDPFARARAVSGPFRF--AFYP 355

DB 324 ADTALHAHYIFNNGIYEETPTDICKEMPMETGMKHTLPRLNALGSIYARAGVIGANVF---381
 DB 356 CYGLAEATLIVSGAREAPV-----TARIAPEEVEELGRAVSAAE--GARFVG 403
 DB 382 -----SNALYILFEVDSJGMEAKDGCPSTNRFYQFAGPPLA-----ANVQIDKGVH 432
 DB 404 SKKALLPKA---VALVDAG-----NELGFG-ELGLTW-VSGISVAAGVQWGRPP-----447
 DB 433 LSSQSTGSSNIERSVDYLALLI--FAPILARILFLFQTA--FTGAGHSLKLVYGT- 490
 DB 448 -----ETEAIFGAILASSAAR-----YLTGIDDELKQGR-----LFVQGRSK 486
 DB 491 -----PDSEIPCSICEKH--TRPVC--AHITVHRIKQM-----PPHQGATIK 529
 DB 487 DLTILGRNHPDPTTEKTESHPAVRPGCSAASFVEHGFEPVAVCEYDPRVADPRE 546
 DB 530 PIGVFCTMNSQYSLQDLDNVAAYLLIKKRGVIFAKAKIMQDLYKATLEKLP--TDLFOER 589
 DB 547 IVAAREAVTAEHQ---LVAHAVALIAPCALPKTSSGKVRHRCRAFLP---DALGR 598
 DB 590 -----LDRGAPPSSEGLSVYVDHPPTPRILTLTPAPTEQTTGFMKVLVETDQK 642
 DB 599 HVAFAPELLDDSPDD---APPETEPGSGRSLILALRS---TIAKALRLDAQDIDAL 651
 DB 643 ---KEGL-SHATHSMALTFUPYSGAPCIHNF 671
 DB 652 PISRFGLDSLAVELQHAFOVRIGRAIPILSIL 684

RESULT 5
 US-09-071-035-358
 Sequence 358: Application US/09071035
 Patent No. 6448043
 GENERAL INFORMATION:
 APPLICANT: Gil H. Choi
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 496
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/071,035
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: A. Anders Brookes
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 358:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1074 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-358

Query Match 2.0% Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3% Pred. No. 0.01;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 240 RYRNTAVAAHAAHAFENHGAVALPPDIITYYFSSSSATTARARRNDVNSTSKPS 309
 DB 67 RTISLYAEYNGAKVIFC-IEPGVSIPTVEVTHGY-----OKNPL 104

QY 410 PSQGFERRIASIM-AAOTLHAEEVIFNIGYEE-----TPTDIKEMPMPI 353
 DB 105 PSMISAKALVSVLMKACGIDIDTNMVAOKMWEYVNGYKXHSIKRIGASVYIK----- 158

QY 454 GMEGLPRNLALGSTAR-----VAQVIG-AMVSPNSALYLTVEVDSGTEKKDGP 405
 DB 159 SLEPKINK-AIEFYQKKPSFHTVTKTILGOSTTILIKNELINSEPKVYVNTA----- 211

QY 406 GPSNRRFQFAPHLAAMPQTDGCHVLSSTSSNTEFSVDYALALCGFAPLAILL 465
 DB 212 NIDRVIGNOLVLP-----NSNSKSGTILTKKSAGTGT-VAAYKK 251

QY 466 FYLERGDAFTGAGDAIKVYVTFEISEIPGSLCEKHTPRVCAHTTVAHRLQRMPPQ 525
 DB 252 AGLOTVMAGALDKPRTVAKINVELTKG-LKIKIDKESGDIPEVTEVL-----DECK 304

QY 526 A-TRQPIQVHCIMNSOVSQDPLQCNVA-----PYLLIKKIQDQFAKAKEM 570
 DB 405 ALPSKQVTTDKKIL-----STIDGIPHTKVITKESVPPPYMIDTPMAATIKAGETI 358

QY 571 QUTYKATLEKIFIDLEJURLDKRACPSSELS-----SVVQDPT-PRILLDTLRA 621
 DB 459 SMISKNNHOKOULLEKRG-VHTGTDIMNDYSLAGNIFALIKRDSQACHVQETIDDKG 417

QY 622 RLEDTTFQFMYVETRYKINEGSEATHSMALTFEP YSAFCEPTINELVKRT 675
 DB 418 KAE TPKELANALELTGYVTE TKSSNFVNTFPTKVELKVAQGVALTSTVNGQ 473

QY 676 HIAVYQIAL-----SQCRCVYQGOV-----EQRNRKRPQVPLARRPVLL 717
 DB 474 NOETIGETTLREKQDTNESGKAFFKGAETTLTAKDGAVKKSEAFK-----TEL 526

QY 718 FNGEISTRTITVILSH-GVSAFNPILAQ-----LAPAGTFEGDGLAKVSE----- 764
 DB 527 VKIILKASLIYALIDQNNVAVANILALINLFWOLNAPKQIILDBITVYSIKKAVINB 586

QY 765 ---VIRDIRKKNRV-----FSGNCTNLSENA-----RRLVGLASAVYQREKR 805
 DB 587 KNAVITRDVTAKEQVIRGPFPEFKNVSAEDTAETGFENLSPKVSPLREITETGAEDKA 646

QY 806 VDMJHGAIGF-----ILKQFHGILPPRCM-----PPNSKS-----PNP 838
 DB 647 TTAVNBOULGFKYGFENLPYGDYILEIEEA-PEGFKITPLEIKSTFEKNKDYAKS 703

QY 819 GMPWTLIAQRNU MPADKLTHEBITTAAVKRFEVEYALINFINLP-PTGIG 888
 DB 704 EYVETTFEEDQKOPIKMIVAYEKLINNE-----FVSINILMLYDIDPEKEDSITS 754

QY 889 LAQFYMANLILAKYCHSOYLINTLISITIGARRPRDPSVLMIRKD-VTSADIETQ 945
 DB 755 LATMKDKNKKIINTIDETE LVQKL-----RYNLHEIKEDWYVAQAADIVEA- 799

QY 946 AKALIEKTENPREMIAFIST-----HIVRAAMNORHVVAGISISKYACA 992
 DB 800 TKAQGERDEKAKPVATATATIANKEKTKWKILHKLITAG-----VLKSIYLFVNV 853

QY 993 ANNNRVQANMSILNCKKNVFLITTEPTREPTIACFEGGCFIPVYTGSSGNRETTISD 1052
 DB 854 YENKVAPEACME-----PVA-----KIASLIN 875

QY 1054 GVRGILVSGAMVQALATVAVAVAGAR-----AGHMAFDMLSTLDEFLARDL- 1102
 DB 876 QAV TYACCTIERIVSIOFKALHEDGSOTFTHGDVMDQFDDVSVTHIDVLD 923

QY 1103 FRLHQLIOLITETFWLVEGALFAVKILDE-----KILAGD-----GEPFINLAFNPD 1149

DB 924 GSKAEVETLIALLLQCLINKLWKSQKIRHYVNDKPRFIVAAKVDCKYTHCTKTF 982
 QY 1150 SCEPSHDTTSNV-----LNISGNSISGTVPGIKRPEDE 1185
 DB 983 -TEIYNEKQNVNCKHNEDLKESQTLTPKEVPTLPSHKQPE 1024

RESULT 6

US-09-071-035-394
 ? Sequence 394, Application US/09071035
 ? Patent No. 6448043
 ? GENERAL INFORMATION:
 ? APPLICANT: Gil H. Choi
 ? TITLE OF INVENTION: Enzyme-catalyzed synthesis of polynucleotides and polypeptides
 ? NUMBER OF SEQUENCES: 496
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Human Genome Sciences, Inc.
 ? STREET: 9410 Key West Avenue
 ? CITY: Rockville
 ? STATE: Maryland
 ? COUNTRY: USA
 ? ZIP: 20850

COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ? COMPUTER: HP Vectra 486/33
 ? OPERATING SYSTEM: MSDOS version 6.2
 ? SOFTWARE: ASCII text

CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/09/071.035
 ? FILING DATE:
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER:
 ? FILING DATE:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: A. Anders Brooks
 ? REGISTRATION NUMBER: 36,373
 ? REFERENCE/DOCKET NUMBER: PB369P2
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (301) 309-8504
 ? TELEFAX: (301) 309-8512
 ? INFORMATION FOR SEQ ID NO: 394:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1074 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? US-09-071-035-394

Query Match 2.0% Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3% Pred. No. 0.01;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RYRNTAVAAHAAHAFENHGAVALPPDIITYYFSSSSATTARARRNDVNSTSKPS 309
 DB 67 RTISLYAEYNGAKVIFC-IEPGVSIPTVEVTHGY-----OKNPL 104

QY 410 PSQGFERRIASIM-AAOTLHAEEVIFNIGYEE-----TPTDIKEMPMPI 353
 DB 105 PSMISAKALVSVLMKACGIDIDTNMVAOKMWEYVNGYKXHSIKRIGASVYIK----- 158

QY 454 GMEGLPRNLALGSTAR-----VAQVIG-AMVSPNSALYLTVEVDSGTEKKDGP 405
 DB 159 SLEPKINK-AIEFYQKKPSFHTVTKTILGOSTTILIKNELINSEPKVYVNTA----- 211

QY 406 GPSNRRFQFAPHLAAMPQTDGCHVLSSTSSNTEFSVDYALALCGFAPLAILL 465
 DB 212 NIDRVIGNOLVLP-----NSNSKSGTILTKKSAGTGT-VAAYKK 251

QY 466 FYLERGDAFTGAGDAIKVYVTFEISEIPGSLCEKHTPRVCAHTTVAHRLQRMPPQ 525

```

Db 252 AGLOTVMAGALDKPNTYAIKINVTGKS-LKTIKKIDKESGDIPTVTFHL-----DFGK 304
OY 526 A-----TRQIGVPGCTMNSOVSICDPLGNVA-----PYLLIRKQGDTEAKAKM 570
Db 305 ALPSKIVTIDKGI-----SLIUKIPIHCTKVITFEKSVDPDPMIDTTPMAATIKAGETI 358
OY 571 QDTYATTELEFTELEOEERILNRCAPCSSEGLS-----SVIYDHT-----FRILDTLRA 621
Db 359 SMISKNMKKQGLIILKIC-VEIGITILMINUNSLACNTFAIRKDSACELVCEITITDEKG 417
OY 622 KIRQITVGMKVLVEIRQKIRKGLSEATHSMALTPDP-----YSCACPTITNPLVKRT 675
Db 418 RAE--TPRELANLELGTYYTE--TKSSNGEVNTEKPTKVELKYANQVVALVTSNVKQ 473
OY 676 HIAVQDIAL-----SOCHCYVYQOV-----EGKNPNQOPVLIRKRPVIL 717
Db 474 NOHITGETITTKKDKPCNSESOKAFKCAEYTLPTAKDQAVKSEBAK-----TEL 526
OY 718 FNGEISTKSIITVLSLSE-SPVSAFNPILQV-----DAPAGTFDGLARVSE-----764
527 VGTASODETVILALDKKNOVAAYKHLAINFYHOGETKAPGCTIDETKYVSVISKVNDNE 586
OY 765 ----VIRDIRKVRNV-----FSGNCTINSEAA-----RAKLVGLASAVOQEK 805
Db 587 KNAVITRDYTAKEQVTFPFTFEKFASSALGTAEFTGNDLSFKVSPLEGTXTITGAEDKA 646
OY 806 VMHICALGF-----LTKOPHGLLPKGM-----PPNSKS-----PNP 838
Db 647 TIACNOLQFDGCKFENLPYGYLLHEIRA---PEGFQKITPLEIRSTFKKKDYAKS 703
OY 839 QMEWTLLOBNQ-----MPADKLTHEITITIAVKKRTEEYAAINEINLP--PTCIGE 888
Db 704 EYVFETTERGQKQPIKMAVIVPYEKITNNF-----FVSILNRLMLYDLPEKEDSLTS 754
OY 889 LQOFYMANILKRCDSOVLNLTLSLITGARRRDPSSVLMIRKQ---VTSADILETQ 945
Db 755 LITWKGNNKLTLDLFE-LVDKL-----RYLHEIKEDMYVAAQIDEA- 799
OY 946 AAALIEKTNLPRLMTATPST-----HIVRAAMNORPVVILGISKYHGA 992
Db 800 IAAQCKDKKAKVIVIAETAILANKKRTGTWKIHKITAEQ-----VLKGSIVLENY 853
OY 993 ANNNVPEVQANMGLNKGKVVPLFTFTTRTRRTIAPPRGFIQPVTPGSSGNRETLSD 1052
Db 854 YENKVAEAGNE-----PVA-----KDAJLNN 875
OY 1053 QVGGIIVSGAMVOLAIVATVRAVGAR-----AQHMAFDDMLSLJDEFLARDL-- 1102
876 QAO-----TNCNTERHVSIGTAKHLEDGSQTFHGVMMFDDVSATHVD 923
OY 1103 --ELHDOIIOTLETPWTEGALBAVKTLDE-----KTTACD-----GETPTNLAFNPD 1149
Db 924 GSRFAETETLYALIPDINKETKSKTEHHVNDKEFTKIVIAKKVUGYPRGCIKIF- 982
OY 1150 SCEPSHDTSNV-----LNSGNSISGSTVPLKRPPEDE 1185
Db 983 -TEINTEKQCNVGNCKHNDIKKESQITTPKEVPTTIPSTPOPE 1024

```

```

: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 5215
: TYPE: PR
: ORGANISM: Streptomyces venezuelae
US-09-105-537-2

Query Match      1.8% Score 111 Db 4 Length 5215:
Best local Similarity 19.7% Pred. No. 3.4 Indels 494 Caps 66:
Matches 250; Conservative 109; Mismatches

OY 51 NLTEKFTFTSSLAVV-----SGAFTTGLAGAGTITKLTTSHPYSPVFPFGSKHLP 102
Db 166 SLTVDAQSSSLVAHYILACESLRAGESTALV-AGVNL-----NILA 206
OY 103 SSAPNLTACNMAARRPFGSCGCPVVDCAVETTCATICTRIGLIPENTILIVIALP 162
Db 207 ESA-----VTEERFG-----GLSPDGTAVTFDARA-- 231
OY 163 KEAVFMGNVFLHVGGLDIYHINHDVIRIPFLFVQLEMPVNRVLPDPENTHHSIGEGF 222
Db 232 -----NCFVRBEG-----GGVAVLKPL-----SRALACDQVH-----CV 261
OY 223 VYTPPEYNTGLCHLHIDCVIAPMAVALPVKNVAVAGAAHLAFDENHGCANVLPDITYT 282
Db 262 IRASAVNNDG-----ATPGLTVPSSRAOGEKVLREAYRRAALDPS-AVQ 303
OY 283 YPOSSSGCTTARGARRNDVNSTSKPSPGCFERRLASIMADPTALHAFVFNCTIVEET 342
Db 304 YVELHGTGT-----PVGDPLEAALGAVLSARADPPL-----LVGSA 342
OY 343 PTDIKEMPEIGMEGTLPRNALGSYARVAVGAVWFSPNSALYTEVEDSCMTEAKD 402
Db 343 KTVNGHLEGAAGIVGLIKTLTALG--RRRIIPASLNFRTPHDPIDP-----DTLGLIVPD 394
OY 403 GGRSPSPNRFYVFAGPLANPQTKRD-----GHVSSQSTSSMTEESVD 448
Db 395 G-----LREMPHPDRELLAGVSSFGMGCTNAHVLSBGPAGGEGEOPRID 438
OY 449 YIALICGPGAPILARILFYLERCDAGA-----FTGGHGDALK-----YVGTGPDSEI-P 496
Db 439 EETPV-----DSGAALPFVVI GRGGEALRAQARLHEAVEADPELAP 480
OY 497 GSLCEK--HTRPVCAHTTV-----HRLPQAMPFPQATRPRTIVFGIMNYSQDCPLG 548
Db 481 AALARSLVTRTPVTHRSVYLAPDRARLDGLALAGTPAPGVVGT----- 528
OY 549 NYAPYLIRKPGDQTEAAKATMDTYRATLERLFTLEOERILDRGAPCSSEGLSVYD 608
Db 529 -----PAG-----RLAVLF-----SSGAGRTGMG-ELYNA 555
OY 609 HPTFRILDTLRLARI EOTTQFMKVLVETRD-----YKIRGLSEATHSMALP 656
Db 556 HPAAITAFDAVAALDPLDRLPRLAELVAAGDTLDRVHTTOPALFAVEVALHRLVESWCVT 615
OY 657 FDPVSGAFPTITNPLVNRHLAVVQDLALSGHVFVGLVWESRNFQPOVLPBRRVD 716
Db 616 PDLLAGH-----SVEISAHYAGV--LSLRQ-----ARLYAAGRLMALP----- 656
OY 717 LENGFISTRSITVLSGVSAPNPITLGDAPAGRTFDGLARVSEVIRDIRKVRNV 776
Db 657 --EGGAM-----VAVEASEEVL---PHL-----AGRERELSLAAN-----GPRAVV 694
OY 777 FSGNCTINSEAKRIRVGLASAYOROEKRVMLHICALCFLIKOPHGLP----- 825
Db 695 LAG-----AERAVLDVAELLREOGRTKRLSVSHA-----FHSPLMEPLDDEFRVV 741
OY 826 -----PR-----GMPNSKSPNPOW-----FWTLLOHNMADKLTHEITITIAV 866
Db 742 ELHDOEHRVIVSVIAGLPVTA-----GQWIDPRYMW-----VQGV-KRPVAFILAV 787
OY 867 KPFTEETAIAINFINLP--TGTIGELAGFYMANLILKYVDHISQYLINTLTITTTAPPPRO 924

```

```

RESULT 7
US-09-105-537-2
: Sequence 2, Application US/09105537A
: Patent No. 6255202
: GENERAL INFORMATION
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438US1
: CURRENT APPLICATION NUMBER: US/09/105.537A
: FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43

```

DB 788 RTLESGAIPLELDHGVCSAMADSV-----KDOEAAVAVASLRKG--RP-E 833
 QY 925 PSSVLIMIRKDVTSAMIEUOKALLEKTENLPELMTTAFSTHLVRAANORPVVLGI 984
 DB 844 POSTLAATTVYRKHEDV-----WTAAKSGTGVVPL---PTVAFOR 874
 QY 985 STSKYGAACNNKHVFOACNNKNGKNCVCLFEPDKTRHLLACDRCGTCGVYCHSSG 1044
 DB 875 ERIMFGKAATAAFLTAGR SGTAG-----TSPAG 905
 QY 1045-----NKETLSDVRCGTVSNGAMQALVATVVA-VIARA 1081
 DB 906 VTSRCHGCHGACAGCDDPAHETT--KRAHAAVAA-----VLKDDPTVVEIGLIF 957
 QY 1082 QJMAFDWMIS-----LTDEFLARDELELHQI10TLEPTWVGALVAVKILDEKTTA 1135
 DB 948 KELGFSLSMVEIENALVDTLKLPLSGILFLP-----PTP---PALAA--HLLSLLTG 1006
 QY 1146 GNEFTTNLAFNDSCGPS--MOTTSNVLNLSGSLVPGCKRPEDDELFDL----- 1189
 DB 1007 GSGCEG--SMAKTPFPAFDLIAEPLALIG---MACRYPGSVISPED--LMRLVAVG 1056
 QY 1190 SCPIKING 1197
 DB 1057 KDAVSGLPIDKRG 1068
 RESULT H
 DB 07 642 734C 4
 Sequence 4, Application US/07642734C
 Patent No. 5824513
 GENERAL INFORMATION:
 APPLICANT: Katz, L.
 APPLICANT: Demadio, S.
 APPLICANT: McAlpine, J. H.
 TITLE OF INVENTION: Recombinant DNA Method for Producing
 TITLE OF INVENTION: Erythromycin Analogs
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Edward H. Gorman
 STREET: Abbott Laboratories D377/ABD-2 One Abbott
 CITY: Abbott Park
 STATE: IL
 COUNTRY: US
 ZIP: 60064 4500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07642734C
 FILING DATE: 17 JAN 91
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Danckerts, Andreas M.
 REGISTRATION NUMBER: 42652
 REFERENCE/DRAWING NUMBER: 4952.US.01
 TELEPHONE: 708 948 2623
 TELEFAX: 708 947 9396
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3567 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DB 07 642 734C 4
 Query Match 1.78: Score 106.5, DB 2: Length 3567,
 Best Local Similarity 19.98: Pred. No. 4.8;
 Matches 237: Conservative 104; Mismatches 384; Indels 465; Gaps 62;

QY 125 CQGPVVDGAVETTGAR--ICGR-----IGLEPENTLXIVVTALEKRAVENCNVFLH 174
 DB 1165 CPGDDLLAAVEEAGASAVVCAQDAALREALGDEP-----VTAL-----VII 1205
 QY 175 YGGDIDVHINHGNYIKLPLPPVQLFMDVNMKLVDPDNTHRHSIGRQVYPTPYNTIGLC 234
 DB 1206 AGTL-----TNGC-----SISVADERP-----AETIAAKTALL 1214
 QY 235 HLHIDCVIAPMAVALRY-----RNVAVAPQAAHL--AFDENEGAVLPDITTT 282
 DB 1235 AVI--DRIVLGHPAPRPVYVSSVAGIWDGACMAVAASAVIALAPRHR----- 1282
 QY 283 YFOSSSGTTTACGARNDNSTSKPSPG-----GF--EKRLASIMADYALHA--EVLPN 335
 DB 1283-----ARGRSTSVAMTPMALPGAVDDGIVREKGLRS--LSADRAMKMEVLA 1330
 QY 336 TGIYEELPDIKEMPMH--GMEGLPLRLNAGSYTAVAGVIGAMVSPNSALYIIEVD 344
 DB 1331 AGHVSAAVADV--DMPVLSRCPAATRP-----TALPAPLACR 1365
 QY 395 SGMTEAKDGSGSPFNFFYFAG-----PHLAANQTDREDGIVLSSQTSSTTE 444
 DB 1366 GGOAEAPRDS--GPTGAPRAQRLAGLSPDQENLLEIVANAVAR-----VLGHESAALINVR 1420
 QY 445 PSVDYIALICGFGAPILARLLFYLERCDAGAPFGCHDALKYVGTPTDSIPGSLCKHT 504
 DB 1421 RARSELGLDLSLNMALRKRL-----SASTG-----LELPASL 1452
 QY 505 RVPYANTTVKTRKQ--HMPPTGALIKVYGVYGLMNSYSLCPL-----GNYA 551
 DB 1453--VFIDHTVTALAOHLKARLVGDAIDQAAVRYGAD--ESHPILAIGICGRPGIGS 1506
 QY 552 PYDLR-----KPGDQTEAA----- 566
 DB 1507 PROLMNVILAGANITTFPAPADRCMDICHLNHPDNDNGTGYVAKGCTTHAADPHDQFPG 1566
 QY 567-----KATMODYRATLERLPIDLEGERLLDKCAPSSRGLSSVIVDPTRK 614
 DB 1567 TTPREAIAMDPOORLMELETAMFAVERAGIDPDAIPGTTTVFVFMNGSQYQGLACEAER 1626
 QY 615 I-----LDITLAKHPQ-----ETQPMKVIYETNDYKIRGLSLATISMAL 655
 DB 1627 VDOYOGIGNSASVLSCHRIATFPMKQIALIVDTACSSSLVCI--HIAOMALRKQCESLAL 1684
 QY 656 I-----FDYSGAFCPITNFIYKRTHLAVYQDIALSOCHVEYGOOVEGRNFEN 704
 DB 1685 AGGVYMSDPYT-----FVDFSTOR-----GLA--SDCKKAFSAADQFALSHQVAAI 1731
 QY 705 QHPYVILKRRFVILFNGHLSIKSLIVLSHQVSAINPILGQAPAG-----RTFKCHLAR 760
 DB 1732 VLEPLSRAR--ANG-----HGVLAVLKGSAAVNGPNSATLAAPNSPGEVIRQALAA 1782
 QY 761 VSEVFLIDILKKNRVVSCNCTNLSFAARARIVGLASAYOROKR-----VDMH-- 810
 DB 1783 SGVPA--ADVIV--VEAHGCTELGDPITAG--ALLATVGDMDKRLIKASVNTINIGHTO 1836
 QY 811-----GALSELLKDFHGLFEPGMPNKSFPQWFTLLQBNQPAOKLTHEETTTA 864
 DB 1837 AAGAGAGVILKVLAMRHMGL--PSLHADDELSPHIDW-----ESGAVVILKEEVWPA 1887
 QY 865 AVKRFTFEYAAINPILNPPICIGFLAOPYMANILKYCHSOVLINWISITIGCARPRD 924
 DB 1888 GER--PPRAGVSSFTVVSST-----NAHIVEEAPAEQ-----EAAETERG 1925
 QY 925 PSSVLIMIRKDVTSAADIEUOKALLEKTENLPEL-----WTTAFSTHHL--VRAMNOR 977
 DB 1926 PLPFLVLSGSRSEAVVAA--QARALAHHLKDTPEHIGITTAAMKTLACRAAPDYKAAV 1978
 QY 978 PMVVLGISTISKYHGAACNNKHVFOACNNKNGKNCVCLFEPDKTRHLLACDRCGTCGVYCHSSG 1037
 DB 1979-----IGDIBASVCAPLDAIARSG-----PSAAVAVAP 2005

OY	1038	VT:PSSNRETTJLWVGLIVS:AMVLATVAVKVRRAUHMFEDELJLTDEEF	1097
		-----VV-----	
DB	2006	VT:-SARPKFVL-----VFPOGQAO-----WVC-----	2026
OY	1038	LARDLEHLLHULLULI-----EIPWIVBGALEAAVILIDERTTAGG	1138
		E:::	
DB	2027	MARPLESEVEFAFSMRCAEALSPHTDW-----KLDD--VWRDGG	2065

RESULT 9
US-08-439-009A-4

GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: McAlpine, J H
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS: Holmstrom, E
CORRESPONDENCE ADDRESS: Holmstrom, E

STREET: Abbott Laboratories CS D377/At6D-2 One Ab
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Parentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A

```

1 CLASSIFICATION: 435
2
3 ATTORNEY/AGENT INFORMATION:
4 NAME: Casuto, Dianne
5 REGISTRATION NUMBER: 40,943
6 REFERENCE/DOCKET NUMBER: 4992-US.D1
7 TELECOMMUNICATION INFORMATION:
8 TELEPHONE: 847-938-3137
9
10 INFORMATION FOR SEQ ID NO: 4:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH: 3567 amino acids
13 TYPE: amino acid
14 TOPOLOGY: linear
15 MOLECULE TYPE: protein
16
17 08-439-009A-4
18
19 Query Match 1.7%; Score 106.5; Dh 3; Length 3567;
20 Best local Similarity 19.9%; Ired. No. 4.8; Mismatches 384; Indels 465; Gaps 62;
21 Matches 237; Conservative 104;

```

```

QY 125 CQRPVGNAGVETIGAE--ICTR-----LGEPPENTILXYTALEKEVAMFNVELH 174
Db 1165 CPHDIIILAAVEACASVVCADDAALREALGDEP-----VTL-----VH 1205
QY 175 YGSLDIYHINBGVDIRLPLEPVVLEMDVDNELVBPDPNTHHRSIGEGFYVDPYNYSLC 234
Db 1206 AGTL-----TFNG-----SISEVADEEF-----AETIAKTALL 1234
QY 235 HLIHDCVLIAMAAVALRV-----RNVTAARGAHL-AEDNEHCVALPDITYT 282
Db 1235 AVL-DEVLGDBRAVEREYCCSVAGICWGACAAVVAAGSYEDALAEHHR----- 1284
QY 283 YFSSSSSGTTAKAKKNDVNSTKPSFS-----GF--ERLASTMADIALHA-EVIFN 335
Db 1283 -----AKRSCISVAMTWALDGCVDDOYTLERGLRS-LSADNKAIRTWERYIA 1330
QY 336 TGIYEETPTDIKEPMFT-ONEGTLPKLNLGSTARVAAGVIGAWFSPNSALTYTEVED 394

```

Db 1331 AGPVSAAVADV-DMPVLSEGFATRP-----TALFAELAGR 1365

395 SGMTEAKDGGPSPFNRFYQFAG-----PHIAANPQIDRDGHVLSQSICSSNTE 444

DL	1366	GGQAEAPDS - CPTGPAQRLAGLSFDEQENLLILVANAFA	VLCBSAAEINVR	1420
----	------	---	--------------	------

445 FSVDTALITGFGAPLLALLFLLEECDTAGAGFTGGHGLALKYVVTTFDSEIPCSEKHT 504

D6	1421	RAFSEGLDLSNMAALRKRL-----SASNG-----LRIPLASI-----	1452
07	505	RVUSADHTTTLVRIYQ--RMYEKGATITROBIVGVEETMNSQVSIDCPRL-----GNVA	551

Db 1453 --VFDDHTVTLAQHLRLRPIVGDADQAAVRVVGAAAD---ESEPFAIVGIGTFPPGIIIS 1506

Q7 552 PYLLR-----KPSDQTEAA----- 566

Db 1507 PEQLMRYVIAEGANLTTGFPADRGWDIGRLYHPDPDNPCISYVDKCGFLTDDAADHPDPCFC 1566

QY 567 -----KATMODYRATLERFLDLEERLLBRGAPCSSEGLSVIVDHPTRRR 614

Db 1567 ITPKRLAMDPOORIMLETTAMEAVEERAGIDPDUALKGTDTGVFCMNGQSYMOLLAGEAFR 1626

OY 615 I-----LEULRARIPT-----TQFMXVIVETRYKIPKGI SEATHSMA, 655
 ; | | | | | ;
1987 INDIAN WIND CHILL RECORD SET BY MCAI DROEGAL 1987

06 656 0-----FIDPPSACAP1TNNFLVKRHLAVVODIALS08CHVFGSOVFNKRN----- 704

Db 1685 AGGVTWSDPYT-----FVEFSTOR-----GLA-SDRCCKAFSKRIDGFALISGVVAL 1733

QY 705 QEPVLRREVDLENGFISTRITVLTSEGPVSAPNPTLGQDAAG---RTFDGLAR 760

Db 1732 VLEPLSRK---ANG-----HQVLAVLRGSAVNODGASGLAAPNGPSQERYIRQALAA 178

QY 761 VSEVIRDIKVNRYVFGSGNCTNLSEARARLVGLASAYQOEKR-----VDMHL-- 810

Db 1783 SCVPA-ADVDV--VEAHGTETELGDPLEAG--ALIATYQODRDRPLRLGSGVKYNIGHIQ 1833

1837 AAGACAGYKVVLMRHGMI-PRSIHADELSPHIDM-----ESGAVFVLRHEVWPWA 188
 811 ----GALGELQFGHLLPRGMPSPKSPNPGFWLIDLRNOMVADRLLHFEITIA B04
 0Y

DZ 1097 MHOHROVYIN VLEBIMORU K NODIRIDDOE LIZON MOCHETIEMOCHENIA
QY 865 AVKRTEEVAINLNPPTIGELAOFYMANLIK YCHISOYLINTLLTSITGARRPRD 924

Db 1888 GER --PRAGVSSFCVSGT-----NAHVIVEAPAPQ-----PAARTERG 1922

925 PSSVLHMIKKDVTSADIEIOKAI,LEKTEINPEL-----WTTAFTSTH,-VRAMNOR 977 QY

Db 1926 PLPFVLSGSEAVVAA---GARALAEHLDTPELGLTDAWTLATGSRAPFNVRAAV--- 197

978 PMVVLGISISKYHGAGNNRFFQAGNWSGLNGCAVNCPLFTYDRTKRFI IACPGCGFICP 103

DB	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
DB	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100

D6	2006 VT---	SAPKPKVL-----VFPGGAGAO--WVG----	202
CY	1036 VLFSSGNELILSDGVNGVLLVSQGGPVVGLLRLTAVNRYGANGVLTUPLCLMNDLALDQLD	:	107
	1038 VLFSSGNELILSDGVNGVLLVSQGGPVVGLLRLTAVNRYGANGVLTUPLCLMNDLALDQLD	:	109

QY 1098 LARDLEELHQI IQTL-----ETPWVEGALFAVKIIDKTFACIGS 1138

```

:||||| : .. ||||| |||
Db 2027 MANDLLESEVFFAESMSRCAEALSPHTTW-----KLTL-VVPIDG 2065

```

RESULT 10

US-07-731-157A-7
: Sequence 7, Application US/07731157A

Patent No. 545/032

GENERAL INFORMATION:

APPLICANT: Quax, Wilhelmus J.

APPLICANT: Misset, Onno

APPLICANT: Van der Laan, Jan M.

APPLICANT: Ientling, Herman B.M.

TITLE OF INVENTION: Mutated beta-lactam acylase genes

```

NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSER: COOLEY GODWARD CASTRO HODDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent to Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,157A
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18 APR 1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBK0-027/0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 494 7622
TELEFAX: 415-857 0663
TELEX: 480816 COOLEY PA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83
US-07-741-157A-7

Query Match
1.7% Score 106; DB 1; Length 774,
Best Local Similarity: 28.2%, Prod. No. 0.31,
Matches: 145; Conserved: 79; Mismatches: 251; Indels: 242; Gaps: 37;

UY 208 AGCAAAATADENHAKAV LPPHLYTYTFOSSSGTTAGAKRNHVNSTSKISPSGCFER 316
DB 184 AANAUKLAKYKXGOLITPCVEAERLEADIALRPVADALIKMGCGASJAAGCGSN 242
317 BLAS---LMAADTALIAEVFTNGIYEETPTDIKEMPMFTMGEGTLPRNALGS 367
241 MAVAVATRTATGAPLACDHPPEPT---PGMYAQHHLACDHPHM-IGL---TV----- 289
UY 408 VYARVACVIGAMVFSTNSA-----LVLFEVDSQMTF----- 399
DB 290 ---GVGPFPEFAHNEVVAVGVTHAFMDLHLLYLEGFAEMRPTAFGPFEPVAMPBD 343
UY 406 AKIKGNHNSNRYGVAFGPHLAAMPQIDRNGHVLSSSTGSSNTPFSVDYIALICGG 457
DB 444 KLAIVKGAIDREDFVTRRGVYIADP---LPCAALILRSVQAFEDLSFDDLTMP--G 398
UY 438 ATLAKLLEFLERLAGATG GH GLAKLVYTGTFDELPCLSLKEKHKIPVLAHT 511
DB 399 ASTVAOLY---DATKWCILIDHNLVAGDVAGSLGHLVYRAVYDRPENCMLVYPCWS 452
UY 512 IVHRLR---GRRHFGCATKOPICVCTMNSQY-----SDCQHLGVAYV--- 553
DB 433 GEHEKGMVPIHEAMP---VIDPFGILVYANNRVVADDPYLCTDCHP---PYRAE 504
UY 554 -LILKKPGIOTFAAKATMOPT---YRAILEKLFI--DLRQEH----- 590
DB 505 KIMKELVAPPAVADIAAIAHATITSPHGLKARLEALGICGSLPAEELKOTLIAMDOR 564
UY 591 LDRGAPSSSELSVAVIHPTEFRILPTLRKILFQTTQFMKVLVETRDYKIREGISEA- 649

```

```

DB 565 MDAGSQAAASA-----YAFRRAL-----TRLTATSGLEQAI 596
UY 650 IHSMALL---PIVYSAGPCH1INFLVKRHLAAVU-----UAIASQCHVYGGVGRNP 702
DB 597 AHPFAVPPGVSPGGVMAVPT-LIRNDAGMLKGSWDEALSEALSV-ATONLTGRGM 654
UY 703 RNOFOPVLRREVDLFGNGFISTRITVTLSEGPVAP-----NFTLRQDAP 749
DB 655 GRHRP---RTHPLLSAOPPAMAALL-----NPVSKPIGGIDGVYLANGVYSAGHPAT 705
UY 750 AGTFEGDLARVSVEVIRIVRNHVFVS--GN-TNLSFAAPLVSLASAYVPEQEKVDM 808
DB 706 YG-----ALSRVFDVGNMND-----SRWV 726
UY 809 LHCALGPHLKQPHGLLPKCMTPNSKSPINQW-----PWTILORNDMPDKLTHIEL 860
DB 727 FICASG-----HPASPHYADQNPWSDCAMVPMLYSMDKILAEAVTSOEL 771

RESULT 11
US-08-541-780-7
Sequence 7, Application US/08541780
Patent No. 5935831
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Missel, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSER: COOLEY GODWARD CASTRO HODDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent to Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBK0-027/0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83
US-08-541-780-7

Query Match
1.7% Score 106; DB 2; Length 774;

```

Best Local Similarity 20.7%; Pred. No. 0.31;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;

```

QY 258 AKGAHAEIENHNGAV-LIPVITYIYVSSSSGTTTAKGAKRNVNISKSPSGGPRK 316
DB 183 AANMLKLEPYDKRQJDLT-LPISVEAEKLEADLAALPRAVDALLKAMGSGASDAAGSNN 242
QY 317 -----RLAS-----TMAADTALHAETIENGTIYEETPTDIKEMPMIGMEGTLPRLNAGS 367
DB 243 WAWAGKRAICGPIIAGIDHKKVYH-L-POYAGHHIACITHTM-IGL-TVP----- 289
QY 368 YTAIVAGVIGAMVESPNSA-----LYLVEEDSGMT----- 399
DB 290 -----GVAFEPFHAKNSKVAVCTHAFMDIHLYLEQFAEDGRTARFANEPEVAMPFD 343
QY 400 ---AKDGGCPSTNRRFYQAGPHLAANPOTDRGCHVLSOSSTGSSSTNTSFTSVYIALICGCG 457
DB 344 RIVAGGADPEFDIETVETHTGPIVAGDP---LEGAALTLSKVGFATDLSFDCGLTTRP--G 398
QY 458 APLARLLEYLERCDAAGFTG--GH-----GDALKYVTGTDESEIPCSEKHTRPVCAHT 511
DB 399 ASTVAQLY-----DATKGWGLIHHNIYAGDVAGSIGHIYKARVSRRENQMLVYPCMS 452
QY 512 TVHRLR-----QKMFPGQATRPGLVFGSTMSQY-----SDCPLGNVAPY--- 553
DB 453 GEHEWRCTIPHEAMPB---VIRPPGGLITVANNRVADDPDYLCIDCHP-----PYRAE 504
QY 554 -----LLKKPGQIIEAAKAIWQI-----YKATLERIFI--DLEQEPH----- 590
DB 505 RIMELVAVSPAFAVDAAIHADTILSPHVGILLARLEALGIGSLPAEELRQTLTAMGDR 564
QY 591 LDPAGPSSSEGLSVYVHPTRFIIDTLRARIEQTTOFMKVLVETPDYKIPRGLSEA- 649
DB 565 MDAGSOAASA-----YNAFRKAL-----TRIYVTRSGLEQI- 596
QY 650 THSMALT---EDPYGACPFITNPLVKRTHLAVO---DLASCHICYFGQOVBGRNF 702
DB 597 AHPEAAVPPGVSPQGVWMAVPI-LLRNDAGMLKMSWDEALSEALSV-ATQNTLTGKWM 654
QY 703 KNOEPVLRIRRVVDLFPNGGFTSTRSITVTLSGPPVAP-----NPTLIGODAP 749
DB 655 GGEHRP---RTIHPLSAQFPAMALL-----NPYSRPIGSGDGYLANGLPSAGREPAT 705
QY 750 AGRTFDGLARVSEVETRIKKNRVFS-IGNTULSEARARLVGLANAYQDQEKPRVM 808
DB 706 YG-----ALSRYVPDVGWMDN-----SRWV 726
QY 809 LHGALGFLIKOFGHLLFPRGMPNSKSPNPOW-----FWTLGONMPADKLTHEE1 860
DB 727 FHGASG-----HPASPHYADONAPMSDCAMVMLTSMRTIAEAVENTSOEL 771

```

RESULT 12
US-08-633-760-46
Sequence 46, Application US/08633760
Patent No. 5804429
GENERAL INFORMATION:
APPLICANT: NIMA, MINEO
APPLICANT: SAITO, YOSHIMASA
APPLICANT: FUJIMURA, TAKAO
APPLICANT: ISHII, YOSHINORI
APPLICANT: NOSHITA, YUTU
TITLE OF INVENTION: A NEW CRYSTALLINE POLYMER
NUMBER OF SEQUENCES: 64
CORRESPONDENT ADDRESS:
ADDRESSEE: ORION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM.

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/633,760
FILING DATE: 01-MAY-1996
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: ORION, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-929-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO. 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-633-760-46

Query Match 1.7%; Score 105; DB 1; Length 774;
Best Local Similarity 20.7%; Pred. No. 0.39;
Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

```

QY 109 LTR--ACNAAREPFGSSRCGPPVCGAVETTGAE-ICTR-----LGLEPENTILYLVTAI 161
DB 70 LTRKRALGRAAEMLG--AEAAEADILVRLGEMKXCRRDFEALGEAVD-----M 117
QY 162 FKEAVPMCVFLHYGGLDIVHINHGDIHRIPLPVOLFMDVRIIVDPPTNTH-----K 216
DB 118 LRAVYAGVNAFLASGA-----PL-PVETGLIGAR---PEVETPHNSIAVAK 159
QY 217 STGEGFVTPPYNTGLCHLHDCYIAPMAVALRPNNTAV---ARGAHLAEFENHES 272
DB 160 RLG-----LIMGAVWFKIMRMIALPVGAANALIKLYDDCGRD 197
QY 273 AV-IIPDITVYVFOSSSGTTARAPRNIVNSTSKSPSGCFEP-----PLAS--IM 322
DB 198 LCIIPPGADRLREADLALTRAVDALLKAMGSDASDAAGSSNNNAVAPGRATATGPIL 257
QY 323 AADTALHAETIENGTIYEETPTDIKEMPMIGMEGTLPRLNAGSYTARVAGVIGAMVS 382
DB 258 AGDPRHVFEL---PKYAGHHIACDRFDM-IGL-TVP-----GVGPRPHFA 298
QY 383 PNSA-----LYLVEEDSGMT-----AKDGGGSPSNRF 412
DB 299 HNGKAVCYTHAFMDIHDLYLEQFAEGRTARFGNDEPEVAMSRDRIAVRGADREPDIV 358
QY 413 YQAGPHLAANPOTDRGCHVLSOSSTGSSSTNTSFTSVYIALICGFGAPILARILFYLRG 472
DB 359 ETRHSGPIVAGDP---FEGAAALTLSKVGFATDLSFDCGLTTRP--GASTVAQLY-----D 407
QY 473 AGAFTG--GH-----GALKYVTGTDESEIPCSEKHTRPVCAHTTVHPIY-----QMP 521
DB 408 ATIRGWLIIHHNIYAGDVAGSIGHIYKARVSRRENQMLPVPMQSGEHHKWM1PHRAMP 467
QY 522 PRQDAROPRIGVFGTMSQY-----SDCPLGNVAPY-----LILKKPDOTE 544
DB 468 K--VIDPFGIIVANNKVVADDPDYLCIDCHP-----PYAEKIMKRIVANIPAAV 519
QY 565 AAKATMONTVPATIFRIETDLRQERLLDPCA--PSSSEGLSSVIV-----D 608
DB 520 DAAAIHADTISP---HVSILRRILEALGARDSSAEGIPQMLVAVDGMADAAEVAASA 574
QY 609 HETFERIIDLTL---RARIQTTOG-----FAKVIYETRDYKIRKG--- 645
DB 575 YNAFRKALIKLYDKRGLDQALISHPFAAIVAGVSPQGVWMAVPIILIKRDDACMLKQMSW 634
QY 646 ---LSEA-----THSMALTDPYSGAGCP 666

```

DB 635 DQALSHALSVASQNLIGRSMGEHRRPRTPIATQPPAMAGILNP 679

RESULT 14

US 08 633 760 48
Sequence 48, Application US/08633760
Patent No. 5804429

GENERAL INFORMATION:

APPLICANT: NIMA, MINO

APPLICANT: SATO, YOSHIMASA

APPLICANT: FUJIMURA, TAKAO

APPLICANT: ISHII, YOSHIMORI

TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: OHION, SEIVAK, McCLELLAND, MAIER & NEUSTADT,

STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/633,760

FILING DATE: 01 MAY 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OHION, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-929-0 PCT

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US 08 633 760 48

Query Match 1.7% Score 105; DB 1; Length 774;

Best Local Similarity 20.7% Pred. No. 0.39; Mismatches 241; Indels 242; Gaps 37;

Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

DB 109 LTR AENAREKEGSPGPPVGAFTTGAETTP---LLEPENTILLYVTAL 161
DB 70 LTRKALDKAAEMLG AFAAENDILVRLGMEKVGCRDEPALKVEKND-----M 117
DB 162 EKAEVMEKNGVLRGGGLDVIHINQDVIKILYFVQLMPVNVNRLVDPPTNTH-----R 216
DB 118 LRAVAVAGNATLSGA-----PL-PVEYGLGAE--PEPWEHISIAVMR 159
DB 217 SEGRFVYPTFEYNTGCHLILHPVIAPMVALVRNTAV---ARKAHLAFENIEG 272
DB 160 KLG-----LMGSAVPMKIMRMALPVCAANALKIRYDGGRD 197
DB 273 AV LPDITTYFQSSSGTTTARGAPPNVNSTSKPSPSTFEF-----RLAS---IM 322
DB 198 LITIPGAEMLKRLADLALTRPVAILLKAGSGDASDAGGSNNWAVAPGTATGRTIL 257
DB 423 AADIALAEVIFNIGVETTPDIKEMPTOMSGTLPRLNALGSIYARVAGVICAMVFS 382
DB 298 AGPGRVFEI---EGIVAGIHLACIDREFOM-IGL--TVP-----GVPFEPHFA 298

QY 383 PNSA-----LYLTVEEDSGMTE-----AKDGGPSPFNRF 412

DB 299 HNGKAVCYTHAFMDIHDLYLRFQAGEGRAPRCNDPEHVAWSRDIIVAGCADREFDIV 358

QY 413 YQAGPHLAANPDQDKGHVLSQSSGSSNTSPSVYIALICGFGAPILARILFYLERGD 472

DB 359 KTRHGVIAGDP---RDGAALILRSVQFAETDLSFCLTRMP--GASTVAOLY-----D 407

QY 473 AGAFPG--GH-----GALKVYVGTPISEIIPCSICEKHTRPVCAHTIVHRIK-----QRM 521

DB 408 AIRGWLIDHNIIVAGDVAGSICHLVVARVPSRPRNGMLVPGWGHMKGWIPHHAMP 467

QY 522 KFSQATKPIGVGTMSQY-----SPCDPFGNVAHY-----LIRKPPDQGE 564

DB 468 R---VIDPFGIILVYANNVAVADHDHYICTDCHP-----PYAKRIKRIYANPAAYD 519

QY 565 AAKAIMQDTRATLERLFIIDLEERILDRCA--PCSSGILSSVIV-----D 608

DB 520 DAAIHADITLSP---HGLLRRLKLEALGARDSDAEGSLRQKLVAMDGRMDASEVANA 574

QY 609 HPTFRILDTL---PARIPOTTQ-----FMKVIYTRDYKIRRG--- 645

DB 575 YNAFRALTRIVTNGSLGLOAISHPIAAVAPGVSPGQYVMAVPTILKDDAGMLKQMSW 634

QY 646 ---LSEA-----THSMALTFDPYSAGFCP 666

DB 635 DQALSHALSVASQNLIGRSMGEHRRPRTPIATQPPAMAGILNP 679

RESULT 14

US-08-931-608A-5

Sequence 5, Application US/08931608A

Patent No. 6302685

GENERAL INFORMATION:

APPLICANT: lobel, Peter

APPLICANT: Steal, David E.

TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND METHODS OF ITS USE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,608A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 601-1-077

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 635 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-931-608A-5

Query Match 1.7% Score 104.5; DB 4; Length 635;

Best Local Similarity 21.0% Pred. No. 0.3;

Matches 77: Conservative 36; Mismatches 121; Indels 133; Gaps 14;

QY 245 MAVALKENVAVAGAAALADENH-EGAVLPULLIITYEYSSSSGILLAKAAHNLVN 303
 Db 10 VAIALAMSSLSAHAD---AWSTHTQAAASPPASTOVLAASTTSATTCNA---YTLN 62
 QY 304 STSKSPSGFERBLSTMAADTALHAETVFTGTIEETPTDIKEMPMFTGMEGLPRIN 363
 Db 63 MGSRLRICA---AVTALKAADHPILHVEVALK-----LRND 95
 QY 364 ALGSTAVACVIGAMVSPNSALVLTVEEDSGMTBAKCGCPSPNNRYOFAPHLAAN 423
 Db 96 ALDTPLAVT-----TPGSALFGKFLTPSQETE-----RF-----GP----- 127
 QY 424 PCHDUGHLSLSSQSSNTEPSVYALICGAGALLAKLFLYLRKQACAGATCGGHA 483
 Db 128 TOSUVAIVAAHLCQAGFTNIEVAPNKL-LISADG-----AGATINCPKIS 172
 QY 484 LKVVLTGTEDELCSLCEKHTPPVVAHTTVHLLKQMPRPBZATRPVIGVGTMSQYSD 543
 Db 173 IK-----RFSANCR----- 181
 QY 544 CIPRLNAYATYLLKRPQCLTAAKALIMUTTRAILPRILPILKQELHKGAMTSSGJLS 603
 Db 182 EFANDAPALVPASTGDSVNAVLLSLQNVSVKILTHIVY---HPECVTVGPVNVITQAAA 237
 QY 604 SVIVDHP 610
 Db 238 AVAAHHP 244

RESULT 15
 US-08-019-870-5

Sequence 5, Application US/08019870

Patent No. 5336613

GENERAL INFORMATION:

APPLICANT: NINA, MINEO

APPLICANT: YOSHIMASA, SAITO

APPLICANT: SASAKI, HITOSHI

APPLICANT: ISHII, YOSHINORI

TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/019,870

FILING DATE: 19930219

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5336613man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-791-0

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-019-870-5

Query Match 17%, Score 104, Db 1, Length 774;

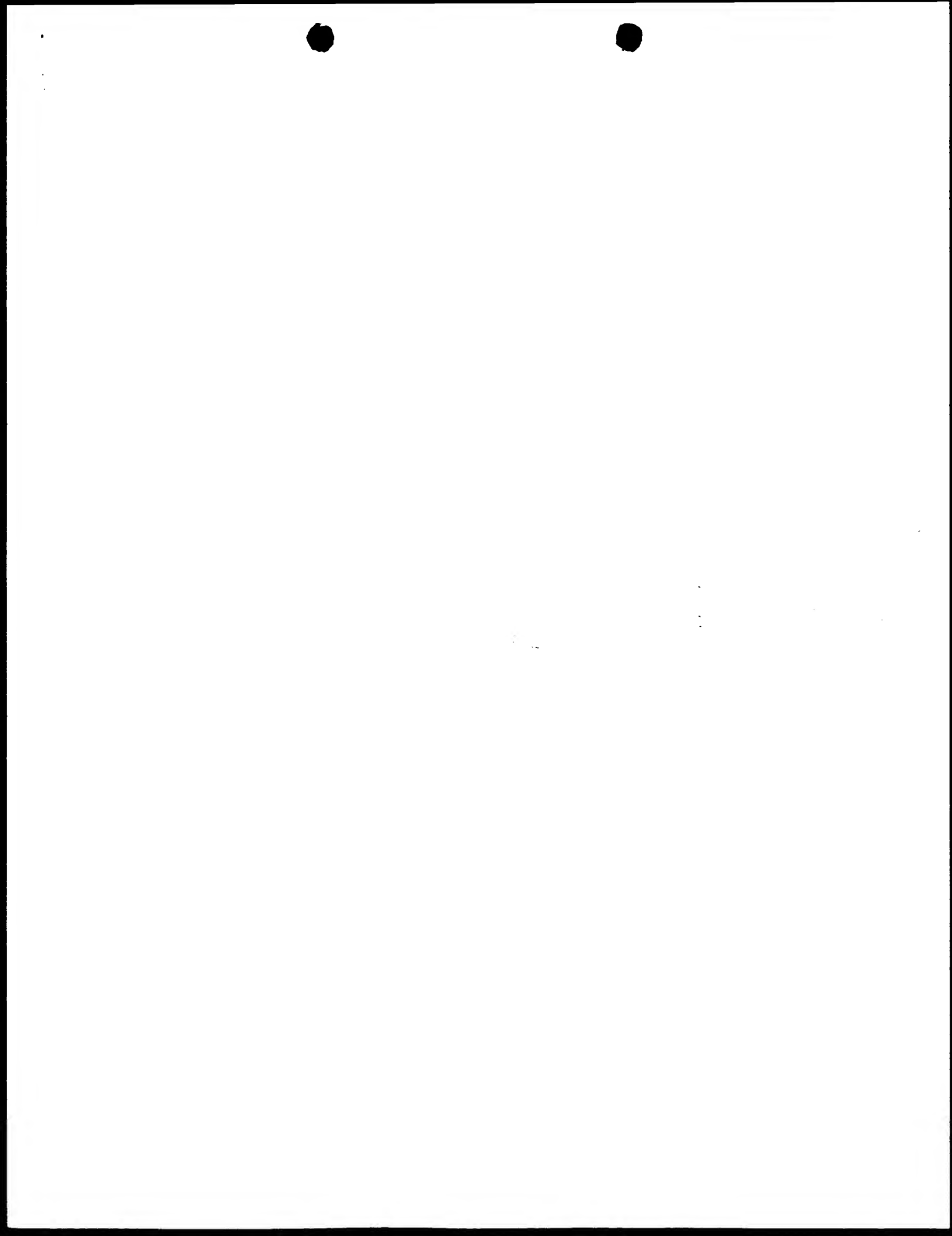
Best Local Similarity 21.7%; Pred No 0.44;

Matches 91; Conservative 41; Mismatches 136; Indels 151; Gaps 19;

QY 372 VAGVIGAMVSPNSA-----LYLVEEDSGMTE-----AK 401
 Db 288 VPGVPGFPRAHNGKAYCVTAHAFMDIHLYLEQFAGDEGTARPGNDPRVAMSRRIAY 347
 QY 402 DGGPGSPNNRYOFACPHLAANDPDRDGHVLSSTGSSNTFSVDYLA-----L 453
 Db 348 KGGADREFPIVETPRHCPVLAGDP---PDSALTLRSVGFETDLSTFTLRMPGASNLV 404
 QY 454 CGPGAPLILAKLFLYLRKQACAGFTGGHGLKVTGPPD-SELPCSLCEKHTPPVCAHTT 512
 Db 405 AGVVAAGSTGHLVTVACLYDA---TRG-----WGLIDRAVPSRPRENGMILPVQWGC 453
 QY 513 VHRLR-----QMPRPBZATRPVIGVGTMSQY-----SL*LPGLANYAPY----- 553
 Db 454 EHEWRCMIPHEAMP---VIDPFGIIVTANNRVVAVDHDDYLCIDCHP-----PYRAER 505
 QY 554 ---LILRKSTGLTAAKATMGLTYKALIEHLFTDIEHKLIMGA--PCSSGJLSVIV- 607
 Db 506 IMKRLVANPAFVAVDDAAAIHADTLSP---HVGSLRFLLEALVARDDSAAERILPMIVA 560
 QY 608 -----DHPFRRLIDTL---RARIQITTO-----FMKVL 634
 Db 561 WDCRMDAASGVASAYNAFRRLTRILVDRGCLDQALSHPPAAVAPGVSYQGYWMAVITL 620
 QY 635 VETRDYKIREG-----LSFA-----THSMALTDPTSGAFCP 666
 Db 621 LRDDDAGMLKQWMDQALSEALSVASONLTGRSGWGERHRRPFTPIATOPPAVAGMLNP 679

Search completed: March 28, 2003, 13:39:45

Job time : 42 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 13:33:34, Search time 54 Seconds
(without alignments)
2141.661 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTOKITVPTGPIKHYVA...

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR-73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Print No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	6294	100.0	1204	1 DNBEC9	DNA-binding protein
2	3620.5	57.5	1209	1 DNBEC4	DNA-binding protein
3	3583	56.9	1208	1 DNBEC74	DNA-binding protein
4	3054	48.5	1136	1 DNBEC5	DNA-binding protein
5	3049	48.4	1136	1 DNBEC1	DNA-binding protein
6	3047	48.4	1136	1 DNBEC1	DNA-binding protein
7	3003	47.7	1197	1 A48350	DNA-binding protein
8	2933	46.6	1186	1 DNBEC6	DNA-binding protein
9	1094.5	17.4	1375	2 J00846	DNA-binding protein
10	669	10.6	1138	1 Q08547	DNA-binding protein
11	633	10.1	1128	2 T13922	DNA-binding protein
12	618	9.8	1145	2 S55600	DNA-binding protein
13	604	9.6	1128	1 DNBEC1	DNA-binding protein
14	596	9.5	1127	2 T03105	DNA-binding protein
15	490.5	7.8	1132	2 T44001	DNA-binding protein
16	483	7.7	1131	2 T41943	DNA-binding protein
17	479.5	7.6	1160	2 A36256	DNA-binding protein
18	477	7.6	1235	1 Q08547	DNA-binding protein
19	412	6.5	1191	1 A44051	DNA-binding protein
20	237.5	3.8	483	2 S69894	DNA-binding protein
21	137	2.2	1504	2 T17426	DNA-binding protein
22	120.5	1.9	781	2 A13014	DNA-binding protein
23	120.5	1.9	788	2 F98269	DNA-binding protein
24	119.5	1.9	781	2 A13014	DNA-binding protein
25	117.5	1.9	781	2 A13014	DNA-binding protein
26	116.5	1.9	781	2 A13014	DNA-binding protein
27	115.5	1.8	735	2 H81702	DNA-binding protein
28	115	1.8	823	2 B35963	DNA-binding protein
29	115	1.8	2505	1 XYRTFA	DNA-binding protein

30	114	1.8	1215	2 152882
31	114	1.8	1941	2 T23979
32	114	1.8	1943	2 T23986
33	113.5	1.8	1886	2 S04921
34	113	1.8	1682	1 C70588
35	112	1.8	735	2 F87355
36	111	1.8	1699	2 T14074
37	111	1.8	3255	2 G81702
38	110.5	1.8	878	2 A87385
39	110.5	1.8	2609	2 T40399
40	110.5	1.8	1069	2 H70656
41	110	1.7	824	2 S24108
42	110	1.7	1039	2 A10122
43	109.5	1.7	977	2 C72526
44	109.5	1.7	1501	2 T45623
45	109.5	1.7	1762	2 T03222

ALIGNMENTS

RESULT 1

DNBEC9

DNA-binding protein - human herpesvirus 3

C:Species: human herpesvirus 3, varicella-zoster virus

C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: C27214

R:Davidson, A.J.; Scott, J.E.

J. Gen. Virol. 67, 1759-1816, 1986

A:Title: The complete DNA sequence of varicella-zoster virus.

A:Reference number: A27345; MUID:86306657; PMID:3018124

A:Accession: C27214

A:Molecule type: DNA

A:Residues: 1-1204 <DAV>

A:Cross-references: EMBL:X04370; NID:q59989; PION:CAA27912.1; PID:q60018

A:Gene: 29

C:Superfamily: herpesvirus DNA-binding protein

C:Keywords: DNA binding

Query Match	100.0%	Score 6294	DB 1	Length 1204
Best local similarity	100.0%	Pred. No. 0		
Matches 1203	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MENTOKITVPTGPGYVACREVDLEDEISFLARSTSDALLPLMFNLTVERKTS	60	
DB	1	MENTOKITVPTGPGYVACREVDLEDEISFLARSTSDALLPLMFNLTVERKTS	60	
QY	61	SLAVVSGARTTGLAGAGITLKTTHSHFPPSVFPHGGRHVLPSSAAPNLTACNAAREPF	120	
DB	61	SLAVVSGARTTGLAGAGITLKTTHSHFPPSVFPHGGRHVLPSSAAPNLTACNAAREPF	120	
QY	121	GFSSRCQPPVGVAVETTGAEICTPLGLEPENTILLYVTALFKEAVFMKNVPLHGGDI	180	
DB	121	GFSSRCQPPVGVAVETTGAEICTPLGLEPENTILLYVTALFKEAVFMKNVPLHGGDI	180	
QY	181	VHINNGVDIRLPLFPVQLFMPDVNRLVDPDNTIHRSIGEGVYPTPYNTGLCHLIDC	240	
DB	181	VHINNGVDIRLPLFPVQLFMPDVNRLVDPDNTIHRSIGEGVYPTPYNTGLCHLIDC	240	
QY	241	VIAVMAVALRVNVTAVARGAHLAFDENHGAVALPDIITYYPOSSSGTTTACGARN	300	
DB	241	VIAVMAVALRVNVTAVARGAHLAFDENHGAVALPDIITYYPOSSSGTTTACGARN	300	
QY	301	DVNSTSKPSPGGEFERRIASMAADTALHAVENTGTIETPTDIKEMPFIGNETLP	360	
DB	301	DVNSTSKPSPGGEFERRIASMAADTALHAVENTGTIETPTDIKEMPFIGNETLP	360	
QY	361	RLNLGSTARVAVGVIGMVPSPNSALYTEVEDSGMTEADGGGSPFNFPYQFAGHL	420	
DB	361	RLNLGSTARVAVGVIGMVPSPNSALYTEVEDSGMTEADGGGSPFNFPYQFAGHL	420	
QY	421	AANDOTDRDCHVLSQSSGTSSNTEFSVDYLAICGFGAPLILAKILFLYLRDCAQATGCH	480	

A:Reference number: A28601; MUID:88179536; PMID:2833010
 A:Accession: A28601
 A:Molecule type: DNA
 A:Residues: 1-1196 <MG>
 A:Cross references: GB:M20165; NID:q340120; PIDN:AAA45793.1; PID:q330121
 C:Genetics:
 A:Map position: 0.38-0.409
 C:Superfamily: herpesvirus DNA-binding protein
 C:Keywords: DNA binding
 Query Match 48.5%; Score 3054; DB 1; Length 1196;
 Host Local Similarity 49.8%; Pred. No. 1.8e-223;
 Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;
 1 MENTOKTAT---VPTGPIGYV--ACRVADLDEISFLAARSTSDALPLMNLIVE 55
 1 METKRTATTTKPPGPIGYVARGPSEGTCL--LALLSARSGADVAVALVGLIVE 58
 56 KFTSSIAVSGARTTGLAGAGITLKLTHSHYPSVVFPHGKHVLPSSAAILNRACNA 115
 59 SCFEANVAVVGSRTTGLTAVSLKLPISHYSSSVYFEGGRILLDPSTQANLRLGER 118
 116 ARERGFESKCGPPVAVAVETTGAEICTRLGLEPENTILYLVVTLFKEAVFMCVFLHY 175
 119 ARRHGFSUDYTPRPDDLKHETTGALCKRLGLDPRKALLYLAVTGPKRAVCIINNTFLHL 178
 176 GLDLIVHINCHDVIRPLEPVULEMPDVNRLVDPFNTIHRSIGEGFYPPPEYNTGLCH 235
 179 GGSCKVTITGAGAVHRIPIYPLDLPDPSKVIAEPFNANHRSIGENFTYPLPFRKPLNR 238
 246 LHHKVALPMAVALKRVNVAVARCAHALAHDEHNEGAVLDPDITVYFQSSSCHTTAR 295
 249 LTFEVAVPAAVAVLGRVNDVAVRAMAHADENIEGALLPADITFEAFASQG--KTPR 296
 296 GARRNDVNSTKPSGSEERELASIMADTALHAEVIFENGIVEETPELIDKEMFIMG 355
 297 GCR-----DCCGKGAGCFKRIASVMAAGDALALHSIVSMAVDPDPPLDLSAMFLCEGQ 351
 456 ECLTPELINALGISTARVAVIGAVFSPNSALYLTLEVEDSGTEAKDGSPGFSENRFOUF 415
 452 DTAARAARVAVAYLAKAGLWAMVSTNSALILETVDACFADPRHSK--PSFYREFLY 410
 416 AGPHLAANPQIDRIKCHV-----SSQSTGSSNTPFESVNTIALICGCGARILAMLLYTL 468
 411 PGHIVAAAPVDKREHIVVGFEGRTAPLVGQTO--EFGAHLAMLCGSPALLAMLEYL 469
 469 EKDCAFTGGHG--DALKYVTGTEDSEIPCSLCEKHTKPVCAHTVHRLRORMPFGCAT 527
 470 EKKKCGVIVCGHEMIVFRVADSNQTDVPCNLCPTDTHACVHTLMLKRAHHKPFASAA 529
 528 KQPIGVETSNMSYSCTOP/INVAAYLLFLKPRGDTAEAKATMDQTYATLEFLTLEQ 587
 540 KGAIGVEGTMSMSYIDVLGNVAEFALKR--ADGSEFARTIMQETVRAATERVAELET 588
 588 EKLIDKRAVCSSEGISVIVHPIFRRIIDPLHMANIDQTIOPMVIYETRHKKIRHGIS 647
 589 LQYVQAVPTAMKLETTITREALHTVANNVROVDEVEDLMNLEVGREKFRDGLG 648
 648 EATHSMALTFDPSGAGPITNLFVKPHTHLAVVQPLALSYTHVFYQVVE--KFNFRNFQ 707
 649 EANHASTITDIPYALCGPDILQILGRSNLAAYQDIALSQCHGVAGGVSCKRRNRQFO 708
 708 PVLRKRRVDFLNGGFTSTPSTVTLSE--PVSAPNPPLTGLQADAPAGPTDGLARVSEVI 766
 709 PVLRKRRVDMENNELSKRTITVALISGMAICAPSLTIGQTAAPASSPEGGVARTLLEGP 768
 767 RDIIVKKNVVSQNGCTINISFAARAVLGLASAYOQKFRVMJHCAIGFLIKQFGILRP 826
 769 KELIVKSRVLELAGASANASEAKARVASLSQAYQKPKRVDILLSPGLKQFIATAIFP 828
 827 PGMPPNSKSPDWFMTLLGNCMPADKLTHEETITIAAVKRTFEVAAINFINLPTCTI 886
 829 NGKIPGNSQNPQWFWALQNPQIPARILSRHDIETIAFKKFSIDYCAINFINLANNV 888

QY 887 GELAQFYMANLLKCYGDSQYLINILITSLITTCARKKRPQSSVLHMKIVLSAADIETVA 946
 DB 889 SELAMTYMANQLIRYCDLSTHYFINITLITLITAGSRPPSVQAAMV--SAQGAILENCA 945
 QY 947 KALLKRENTLPELMTTAPSTHLVRAAMQRPVVVIGISISYHGAAGNNRPVQAGNMG 1006
 DB 946 RALMDAVDAHPPAMTSMFASCKLLRPVMAARVMYVIGLISIKTYGACANDRPVQAGNMA 1005
 QY 1007 LMGKNVCPLETFDPTPRFTIACPPRGFTCPVTGPSSGNRETTLSQVYGVITVSQAMV 1066
 DB 1006 LMGGNACPLLEFDRTRKVLACPRAGFCVCAASNLGCAHESSLCEQJLITSEQAAVA 1065
 QY 1067 LAIVAVVAVAGARQAHAFDMQLITDEPLARLDELHDLQIQTLETPVWVEKAL--- 1123
 DB 1066 SSVFAITYKSLSPRTQVQLEDMALLLEDEYISEEMELTAKALEKNGEMSTDALEVA 1125
 QY 1124 -EAVKILDEKTYAGCGEPTNLAFNPD--SCPSHDTTNSVNIJSGSNISGSTVQILKRP 1180
 DB 1126 HEAFALVSQLACAGE-----VFNPQVPCGR-----IDNATPPQGPAPAPAGRKRA 1173
 QY 1181 PEDDELFDLSGIPIRKGNITTEM 1203
 DB 1174 FHGDUPFG-KSPDPRKKGDLFLDM 1195

RESULT 5

DNBEV1
 major DNA-binding protein UL29 - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1
 A:Note: host Homo sapiens (man)
 C:date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000
 C:Accession: A03790; B30085
 R:Quinn, J.P.; McGeoch, D.T.
 Nucleic Acids Res. 13, 8143-8163, 1985
 A:title: DNA sequence of the region in the genome of herpes simplex virus type 1 cont
 A:Reference number: A93601; MUID:86067223; PMID:2993714
 A:Accession: A03790
 A:Molecule type: DNA
 A:Residues: 1-1196 <MG>
 A:Cross references: GB:X03181; GB:M12356; NID:q59862; PIDN:GAA6640.1; PID:q59863
 A:Experimental source: strain 17
 R:McGeoch, D.J.; Daillymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D., p
 J. Gen. Virol. 69, 1531-1574, 1988
 A:title: The complete DNA sequence of the long unique region in the genome of herpes
 A:Reference number: A30083; MUID:88274327; PMID:2839594
 A:Accession: B30085
 A:status: nucleic acid sequence not shown; translation not shown
 A:molecule type: DNA
 A:Residues: 1-1196 <MG>
 A:Cross references: GB:D10879; NID:q221721; PIDN:BAAN1675.1; PID:q221720; GB:D00417
 C:Genetics:
 A:Gene: UL29
 A:Map position: 0.38-0.409
 C:Superfamily: herpesvirus DNA-binding protein
 C:Keywords: DNA binding

Query Match 48.4%; Score 3049; DB 1; Length 1196;
 Best Local Similarity 49.8%; Pred. No. 4.4e-223;
 Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;
 1 MENTOKTAT---VPTGPIGYV--ACRVADLDEISFLAARSTSDALPLMNLIVE 55
 1 METKRTATTTKPPGPIGYVARGPSEGTCL--LALLSARSGADVAVALVGLIVE 58
 56 KFTSSIAVSGARTTGLAGAGITLKLTHSHYPSVVFPHGKHVLPSSAAILNRACNA 115
 59 SCFEANVAVVGSRTTGLTAVSLKLPISHYSSSVYFPHGKHVLPSTQANLRLGER 118
 116 ARERGFESKCGPPVAVAVETTGAEICTRLGLEPENTILYLVVTLFKEAVFMCVFLHY 175
 119 ARRHGFSUDYTPRPDDLKHETTGALCKRLGLDPRKALLYLAVTGPKRAVCIINNTFLHL 178

```

QY 176 GOLDIVHINHGIVIRIPLEPVGLFMDLVNR.LVDPFNTNHRIS:IGESFVYPTPEYNTGLCH 235
DB 179 GGSQKAVTIGCAIVHRIIPVYPIQI.LFMPDRSRVIAEPFNANHRISIGEFYPIPLFFPNRPLNR 238
QY 236 L.IHDCVIAPMVAVALRVNVTAVARCAAH.IAPDHNHGAVLPPDITYTYVOSSSGCTTTAR 295
DB 239 L.IFEAVVPAVAALRVNDVAFAAH.IAFDENHGALPADIPTFAEASOG--KTRR 296
QY 296 GARRNDVNSTKPSPSGGERIRLASIMADPTALHAEVINTGICEPTPEPTIKEMPIEIOG 355
DB 297 GCR-----DGGCKGAAGCEQRLASVMAQDALALEISVMAVFDPEPDDISAMP.LFEQ 351
QY 356 EGTLEPLNALGSYIARVAVIGAVVSPNSALYTEVEDSGMTEAKDGSGPSNFRYOF 415
DB 352 DTAARAANAAGAYLARAAGLVGAMVFSTNSALHLETVDAGAPDPDHSK--PSYRFFLY 410
QY 416 AGPHIAANPQTDRIQHV.L-----SSQSTGSSNTRFESVDYIALICGFCAPILARILLYL 468
DB 411 PGIHVAANPVUREGHVYVGFEGRETPALVGGTV--EFAGEHLAMLCGFSPALAKMLFYL 469
QY 469 EPCFAGAFGCHG--DAIKVCTGFSEISLCEKHITPVCATHTYHRLKRMH-FPGQAI 527
DB 470 HKDCAVIVGROEMVYRVVAUSNOTDVPCNICTPDRACVHTILMKRAHPRKASAA 529
QY 528 KQFIVGFTMNSVYSDCEFLGNVAPYLLIRKPDQTEAKKATMODTYRATLERFLDLEQ 587
DB 530 KQALICVPGIMSNYSDDCVLIGNVAAFSALKR--ADGSETARTIMQETRYAATERVMALEFI 588
QY 588 KHLIRHCAPOSSFGSLSVIVHPTFRKILDTPLARILEOTTQPMKVTIVTRYKI.PEGIS 647
DB 583 LQYVQAVPTAMGRLETTITNREALHTVYNNV.KUYVDREVQLMRVLVGRNKFEDSLA 648
QY 648 EATHSMALTFEDYSGAFCHITNFIWKRIHLAVVQDALASQCHVYGVQVVEGKRNFRQFO 707
DB 649 EANHMSITLIDRYACGPCLLIQLGRKSNLAVYODLALSQCHGVFGQSGVEGRNFRQFO 708
QY 708 PYLRRFVDLEFGGFTISTITVTISEG--PVSAFNPITLQODAPAGTPEGDLARVSEVI 766
DB 709 PYLRRVMDMFNGFLSAKTLTVALISEGAALCAPSLTAGOTAPAESSEFGDAVAVTLGFP 768
QY 767 KHLIVKNNVHVSQNCCTNI.SFAAARAVI.VGIAVYQROEKFVDMHGAICPLIKOFHCLP 826
DB 769 KELVAKSPVLFNGASANNSSEAAKARVASTLSAYQKDKRVDLLGFLGLKQFHAIIFP 828
QY 827 KCMPTNSSTNQWTLILGNOMIADKLTHERITTLAAVKRTEFYAALNFINLPYCI 886
DB 829 NGKPGOSNQPNQWTLALORNOI.PARILISRDIETITAFIKKFSIDYGAALNFINLAPNV 888
QY 887 GELAGFYMANLILKYGCHDSOYLINTLISITIGARPPRDSVYIHWIRKQVTSADILEQA 946
DB 889 SHIAATYMANOILKYGCHDSYFINTLITALLAGSRPPSVQAAAM--SAQCGAGLEGA 945
QY 947 KALIKRTNIPELMTAFTSTHIVRAANORPMVVLGISISYHGAAGNNRYFOACNMGC 1006-
DB 946 PALMDAVAIHPACAMISFASCNLDPVMAAPVYVLSLTSYCYGAAQMDPVQAAANMS 1005
QY 1007 LMGCGNACGLLFDRTKRKVLVACPRAGFCVCAASSLQCGHESLCEQLMGIIISEGAAVA 1065
DB 1006 LMGCGNACGLLFDRTKRKVLVACPRAGFCVCAASSLQCGHESLCEQLMGIIISEGAAVA 1065
QY 1067 LAIVATVAVAGARQIHMFDOMLSTLDEFLAROLEELHDQILQTLFPMYTEAL--- 1123
DB 1066 SSVFAIVAVSLGPKVUQIQLHMLALLHDEYLSSEKMEITAAALEGNQPMSTDAALIPA 1125
QY 1124 -EAVKILDEKTTACDGETTINLAFNFD--SCPSHDUTTSNVLINISGSI.SGSTVPIKRP 1180
DB 1126 HEAEVLEVSQSNAGE-----VFNEGDFGE-----DONATFEGSGAGVAFAGRKRA 1173
QY 1181 PEUDLPIPLISGLPIKHQNTIMEM 1203
DB 1174 FHGDPPFC--EGPPDKKGLDILDM 1195

```

RESULT 6
DDBEIR
DNA-binding protein - human herpesvirus 1 (strain F)
C:Species: human herpesvirus 1
A:Note: host Homo sapiens (man)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #extl_change 07-Jun-1996
C:Accession: D29242
R:Hamerschmidt, W.; Conraths, F.; Mankeitz, J.; Paul, G.; Ludwig, H.; Buhk, H.-J.
Virology 165, 388-405, 1988
A:Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus
A:Reference number: A94381; M01D:85306231; PMID:2841793
A:Accession: D29242
A:Molecule type: DNA
A:Residues: 1-1196 <HAM>
A:Cross-references: GR:M21629
C:Superfamily: herpesvirus DNA-binding protein
C:Keywords: DNA binding

Query Match 48.4%; Score 3047; DB 1; Length 1196;
Best local similarity 49.8%; Pred. No. 5,3e-223;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

```

QY 1 MENTQKTV---VPIGPIGVY--ACRVEDLDLEISPLAARSTSDIALPLMRNITVH 55
DB 1 METPKTATTIKVPSPGLGVYVARACPSSEGL--LALLSARSGSDVAVAPLVGLIVE 58
QY 56 KPTSSLAIVSGARTGLAGAGITLKTSHFVSVFVPHGCHVLPSSAAPMLTRACNA 115
DB 59 SCFPAVAVAVVCSPTTICLTAAVSLKITTSHSSSVYVPHGCHILNPTQAPMLTRICER 118
QY 116 ARERFPSRQUGPVDVAVETTGAETICTRLGEPENTILYLVATLKEAVFMCNFIAY 175
DB 119 AKHNFGRSDYTPRPGDLKHETGEALCERGLDPRALILYLVTEBGRKVASINNTPLHL 178
QY 176 GOLDIVHINHGIVIRIPLEPVGLFMDLVNR.LVDPFNTNHRIS:IGESFVYPTPEYNTGLCH 235
DB 179 GGSQKAVTIGCAIVHRIIPVYPIQI.LFMPDRSRVIAEPFNANHRISIGEFYPIPLFFPNRPLNR 238
QY 236 L.IHDCVIAPMVAVALRVNVTAVARCAAH.IAPDHNHGAVLPPDITYTYVOSSSGCTTTAR 295
DB 239 L.IFEAVVPAVAALRVNDVAFAAH.IAFDENHGALPADIPTFAEASOG--KTRR 296
QY 296 GARRNDVNSTKPSPSGGERIRLASIMADPTALHAEVINTGICEPTPEPTIKEMPIEIOG 355
DB 297 GCR-----DGGCKGAAGCEQRLASVMAQDALALEISVMAVFDPEPDDISAMP.LFEQ 351
QY 356 EGTLEPLNALGSYIARVAVIGAVVSPNSALYTEVEDSGMTEAKDGSGPSNFRYOF 415
DB 352 DTAARAANAAGAYLARAAGLVGAMVFSTNSALHLETVDAGAPDPDHSK--PSYRFFLY 410
QY 416 AGPHIAANPQTDRIQHV.L-----SSQSTGSSNTRFESVDYIALICGFCAPILARILLYL 468
DB 411 PGIHVAANPVUREGHVYVGFEGRETPALVGGTV--EFAGEHLAMLCGFSPALAKMLFYL 469
QY 469 EPCFAGAFGCHG--DAIKVCTGFSEISLCEKHITPVCATHTYHRLKRMH-FPGQAI 527
DB 470 HKDCAVIVGROEMVYRVVAUSNOTDVPCNICTPDRACVHTILMKRAHPRKASAA 529
QY 528 KQFIVGFTMNSVYSDCEFLGNVAPYLLIRKPDQTEAKKATMODTYRATLERFLDLEQ 587
DB 530 KQALICVPGIMSNYSDDCVLIGNVAAFSALKR--ADGSETARTIMQETRYAATERVMALEFI 588
QY 588 ERLIDGAPCSSGLSSVYDHPTRRILDTPLARILEOTTQPMKVTIVTRYKI.PEGIS 647
DB 589 LQYVQAVPTAMGRLETTITNREALHTVYNNV.KUYVDREVQLMRVLVGRNKFEDSLA 648
QY 648 EATHSMALTFEDYSGAFCHITNFIWKRIHLAVVQDALASQCHVYGVQVVEGKRNFRQFO 707
DB 649 EANHMSITLIDRYACGPCLLIQLGRKSNLAVYODLALSQCHGVFGQSGVEGRNFRQFO 708
QY 708 PYLRRFVDLEFGGFTISTITVTISEG--PVSAFNPITLQODAPAGTPEGDLARVSEVI 766
DB 709 PYLRRVMDMFNGFLSAKTLTVALISEGAALCAPSLTAGOTAPAESSEFGDAVAVTLGFP 768

```

```

QY 767 RDRVKNVSENGNTNLSEAAAPLVGLASAVQROEKPVMDLHGALLKOPHGLLP 826
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 769 KRLKRVSRVPLACASANASNAKAKVASIQSAVQKQPKKVDILLCPILKQFHAAIFP 828
QY 827 KMPNPKSPNPQWMTLLQRONMADKLTHEITTTAAVKRETEEVAINFNLPTPTCI 886
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 829 NKRPQSPNPQWMTLLQRONMADKLTHEITTTAAVKRETEEVAINFNLPTPTCI 888
QY 887 GELAAQVYMANLLKCYGDHSDYLLNTLTSITTCARRKRDSSVLAHWIKQVTSADILFTQA 946
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 889 SELAMTYMANQILKCYGDHSDYLLNTLTSITTCARRKRDSSVLAHWIKQVTSADILFTQA 945
QY 947 KALLEETENLPMLTFTSTLVLRAAMNORPVVVLGISIKYTGAGANNRVQAGNMSG 1006
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 949 RAIVAVADHAPQAMISPMFASCNILRVYMAARHVVIGLSIKYTGAGANNRVQAGNMSG 1005
QY 1007 LKQKRVNPLFTTDRTRKFIAPPRGAFITPVTPGSSGNRETTLSQVQGITVSGGAMVQ 1066
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1009 LKQKRVNPLFTTDRTRKFIAPPRGAFITPVTPGSSGNRETTLSQVQGITVSGGAMVQ 1065
QY 1067 LAIVATVAVAKAQAQMAAPDWLSITDDEFLARDLLELDHDIQITLFTPWVEGAL--- 1123
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1069 SSVFVATVSSLRGRTQQLQLEMLLDEDEYLSSEEMELTARLENGENKSTDAALEVA 1125
QY 1124 -EAVKILDEKTTAGIDETPTNLAENFD--SCEPSHDTTSNVLNISGSNISTGTVPGILKP 1180
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1126 HAAKAVVSGQGNAGF-----VFNEGPGGPF-----DINATPFGPGCAVPAFACKKRA 1173
QY 1181 PEDELEFDSIGPIKHNITMEM 1203
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1174 FKGDDPGF-EGPDRKKGLTLM 1195

RESULT 7
A48350
DNA-binding protein - human herpesvirus 2
C:Species: human herpesvirus 2
A:Note: host Homo sapiens (man)
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-May-1996
C:Accession: A48350
R:Toh, Y.; Liu, Y.; Tanaka, S.; Mori, R.
Arch. Virol. 129, 183-196, 1993
A:Title: Nucleotide sequence of the major DNA-binding protein gene of herpes simplex virus
A:Accession: A48350; M010:9328441; PMID:8185914
A:Residues: 1-1197 (T=8).
A:Note: sequence extracted from NCBI backbone (NCBIN.129069, NCBI.P.129070)
Genetics:
Map position: 0.475-0.405
Suprafamily: herpesvirus DNA-binding protein
Keywords: DNA binding; zinc finger
E:499-512/Region: zinc finger

Query Match 47.7% Score 3003; DB 1; Length 1197;
Host Local Similarity 48.9%; Pred No 1 Acc 219;
Matches 594; Conservative 212; Mismatches 369; Indels 40; Gaps 16;

QY 4 TQKTVTPPTGATVYV ACPEVDLDFETSLAASSTPSIALTLMLNLTVEKFTSS 61
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 7 TTTTAKVPKPMYVYGRAPAEGLFL--LSLSASGADAVAPLLIGLTVESGFEN 64
QY 62 LAVVGAKRTTGACATITKTTSHTYISVYVHGSKHVLPSAAINLTACAAAREPG 121
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 65 VAAVVGSGTGTGTAVALSLKMSHSVSVYVHGGRLLAPSTQAPNLTPLGERARRHG 124
QY 122 FSKCGGPPVQCAVETGATGCTRLGLEPENTILYLVATLFEAVMVCNVPFLHYGGGLDV 181
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 125 FSDVAPKCHDILKHTTGALCHKICIDIPKALILVITTCFRAVCLSNPLHLGOMDKV 184
QY 182 HINIKGVIRLPLEPVQLMPDPVNRLEVPDPNTHIRSTEAFVYPTPEFNTSLTHLTHKCV 241
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db 185 TIGDAEVHPIPVYPLQMFDPSPVINDPFCNINRSTGENFNPLPEFNPLAILEV 244
QY 242 IAPMAVALRVNRTAARCAHLAFDENHCAVLPDITVTVYQSSSGCTTTAARARN 301
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 245 VCPAAVALAKARNDAVARAAHLAFDENHCAVLPDITVTVYQSSSGCTTTAARARN 299
QY 302 VNSTKSPSGGFEERLRSIMADTALHAEVIFNTGTYEETPDIDIKEMPEFICEGLPR 361
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 300 --DAGKNGPAGCFEQRILASVMAGDALALEISVMAVFEDEPHDITTTMLKQETTAAR 357
QY 362 LNALGSTATVAVGIVGAVPSPNSAIVLTVEDSCMTAAKDGQSPNRPYQAFPHLA 421
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 358 AGAVAVILARAAGIVGAVMSTNSALHLEVDAGPADPDHKK--PSFYRPLVPTHVA 416
QY 422 ANQPPDRDGHV-----SSQSTGSSNTPSPVDYIALICGFCARILARLFTLEKIDAG 474
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 417 ANQPLDREGHVAVQYDGRPTAPLVGGIV--FACGHLAMLCGSPALLAKMLFYLEKIDG 475
QY 475 AFTGHLG-DALKYVTGTFEDELFCSEKHTFVCAHTVHPLRPPMPREGQATPQFV 533
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 476 VIQGRQEMDVFRVYVADSGQTDVPCNICITPFRACATHTLMRLKRRHPRASAARCAICV 535
QY 534 FGTMSQSYSCDPLQNTAPPLILKRGQDTAAKAKIMQDIYKATIERLFTLEKQRLDR 593
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 536 FGTMSASVSDCVLGNVAAPSALKR--ADGSENTRTIMQETVRAATERVMALELALGVQ 594
QY 594 GATCSSRGSLSVIVDHTFRRILDTLRLRIEQTTQPMKVLVETRRYKIREGISPAATHSM 653
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 595 AVPTALGRLEETIIGNRALHTVYNNIKOLVDREVQIMRLIKGRNKFQDLAEANHAM 654
QY 654 ALTFDPSGAFPCPTNLELVKFTHLAVVDLALSLQH--VEFYQGVVEFNPVNPQVPLRR 713
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 655 SLSDPYTCGPCLDILQILARRSNILAVYQDLALISQCHGVAGQGVGGRNFPNGQVILRR 714
QY 714 FVILFNGCFISTRTITVTYSEG--PVGAPNPTIGODAPAGTTRGDIARVSVHTRDILRVK 772
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 715 VMDLENNGLFSAKTLTVALESGAICAPSLTAGOTAPAESSEFGDAVRVLGPEKELRVK 774
QY 773 NRVFSGNCTINLSAARARIVGLASAVQROEKRVMDLHGALFCILKOPHGLLPPRCMPN 832
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 775 SRVLPAFASACASEAKARAVASIQSAVQKQPKKVDILLCPILKQFHAAVITPPNCKPFG 834
QY 833 SKSPNPQWMTLLQRONMADKLTHEITTTAAVKRETEEVAINFNLPTPTCI 892
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 835 SNQPNQWMTALQRONMADKLTHEITTTAAVKRETEEVAINFNLPTPTPTCI 894
QY 893 YMANLLKCYGDHSDYLLNTLTSITTCARRKRDSSVLAHWIKQVTSADILFTQA 952
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 895 YMANQILKCYGDHSDYLLNTLTSITTCARRKRDSSVLAHWIKQVTSADILFTQA 951
QY 953 TENLPMLTFTSTLVLRAAMNORPVVVLGISIKYTGAGANNRVQAGNMSG 1012
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 952 IDAHNPGAMTSMFASCNILRVYMAARHVVIGLSIKYTGAGANNRVQAGNMSG 1011
QY 1013 VCPLEFTEFDRTRKFIAPPRGAFITPVTPGSSGNRETTLSQVQGITVSGGAMVQ 1072
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1012 ACPLILFDTRKRVLAACPAFCVCAASSIGGGAHESHLCPQLKCI LAEGCAVAASSVFA 1071
QY 1073 VRAVGAQAQMAAPDWLSITDDEFLARDLLELDHDIQITLFTPWVEGAL---EAVK 1128
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1072 TVKSLSPRQVQIDEDMLLDEDEYLSSEEMELTARLENGENKSTDAALEVALEVAL 1131
QY 1129 LDEKTTAGCGEPTTNLAFNEDSCPESHDITSNVLANISGSNISTGTVPGILKP 1188
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1132 VSOIGAAAGF-----VFNEGPGGDEDDHAASPGGLAAA--AAGAAVARRRAARHGDIDPG 1183
QY 1189 LSGIPKHNITMEM 1203
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1184 -KGPPEKK-DITLDM 1196

RESULT 8
DNBEGB

```



```

Db 18 PVGACGYIYVYPKGCFPEQFASLIGNRWGADAMSLPLLSGLTVFANPSFNKAVKHKID 77
QY 71 TGLAGAGITLLTSHPEPSVYFVHGKHVLPSSAAPULTACNAREGFESKQCPV 130
Db 78 M-----TTLSVRVSAHREAIIVEFNTEFTPIFFGPGLDLSGDARMLFGTSTP-PRH 130
QY 131 DGA-----VETTGAEICRIAGLEPPNTIIVLTALFKEAVMCMVFLHYCGDILVHMHG 186
Db 131 DLKDLIDIKDLAFETYK-----DSCFMSVVTSESEKRLFGNLYPLISGOKQVJNGR 185
QY 187 DVIRIPFPVQFMPDVNRKLVPPDPNTIHRSIGEGFYPTFPYNTGCLHIDCVIAPMA 246
Db 186 EAVKIPLYDEDLFSKSHRL-----PPFYIPSVSKYLDLSLFTSIA 226
QY 247 VALRVNVTAVARGAHLAFDENEGAVLPDITYTFQSSSSGTTARGRANDVNTS 306
Db 227 QALIRIVESTIRALEQSHIDQYKAKV-----VNSK- 259
QY 307 KPSFSGCHERLASIMADTA-----LAHEVIFNTGI-VEETP-----TDIKEMPMF 352
Db 260 -----EFALQAVKCODASAFVVIDCIAELAISYGLSFLFEPQECALLDYTSWEIF 311
QY 353 IGMETLPLRNALGSYTAARVAVIGAVFSPNSALYLTVEVDSGMEKDGPGPSFNRF 412
Db 312 DITETEGRIKAIQDMAMMSVHYTHLFTNSVLYLTKLNKOTQNKSDN---VNSY 368
QY 413 YQFAGPLAANPOTDRGHVLSQSTSSNTFSDYALICGAPLLARLLELEKCD 472
Db 369 FMQHGISTAAEATQKEMDPAFSGAVKFGSGSYTLFHLALASSPSPHLLANCGYMQDCQ 428
QY 473 AGAFQCHGALAVGTGTPSEIPCSICEKHTRYVCANTYHRLKRMKPRGQATROPIC 532
Db 429 HOKSTINSNTSVMOYVITAVSELCELCQGPCPAACHTHTFRLKDFRPVLLSQRRDPY 488
QY 533 VFTSMNSYSDCDPLGNVAYLILRKPGD-----QTEAAKAIMDYATLIERLIDLEJE 588
Db 489 VAVGVSQVYNDLMLGNFAF---REKEDDMVONTCEKTYMQLIONVYKL----- 538
QY 589 RLLDKAPRCSSEGL---SSVVDHTEFRRLDYLARIEQTTQPMKVLVETRDYKIREGL 646
Db 539 -----ATIGITEGLGDLITDIQSFLEKREIDNVVDNEVIKFQINCLV-NNINFETI 592
QY 647 SEATSMALTFDYSACGPTNPLVYKRTHLAVYODLALSGCHVYVYQ----- 695
Db 593 KSVHHLHGYCNGFWQACMFLNLFKSVLAILIODICLV-AMITEODNPSGMPSE 650
QY 696 -----QVBSRNFN-----QFQVLRFRFYDLFNGGFISTRSITVTLSEGPVS 738
Db 651 WLKMHYQTIWTFNKSSCIDKGVLTGAHKKVYHGMPCDPIINV-----S 694
QY 739 APRTTIGODAPARTITDGLARVSEVIRIDIRKKNRVESGNTNLSEAPARLVGLASA 798
Db 695 ALN---GQIVPV---KMOVRLKALITVPKTIKIKNRIVFSN---SSMEALVQCFI----- 742
QY 799 YOROKERVLD--MLHGAIGLILKQFHGLFPRGMPNPKSPQMFV--TLQORONMPA-DK 854
Db 743 ---SSSTKQDSYITGYMKFLNSLHKVIFP-----DAKSALYMHTEFSQKQRIPLRG 794
QY 855 LTHEETTTA-AKRTTEEAALNFINLPPTCIGELAQYMANLLKYCHSYYLLNTL- 912
Db 795 ISKENIIEIANYIEAGSKMDMNVLDIIPPTLLAAVAKVRLNNTLIFTCGOTOFAYATLQ 854
QY 913 -----TSIITGARPRDPSVVLHMKRDVTSADIEQAK---ALLEKTENLELMWTAF 964
Db 855 CLLPITQVYSAITEPH-----VLH---QOOSITSVDYDLSSTINKQALLIYOTLLKEIDATIG- 907
QY 965 TSHILVRAAMNORPVVGLISISKYHCAAGNNVFOAGN---WSGNGCKNVCPLFT-PD 1020
Db 908 -----KHPIYTVPLVYKRYTGINGNTQIFQCGNLGYFMGRGVDRNLIPDSSGFR 957
QY 1021 RRRRFITACPRGEGICPVG---FSSGNET-----TISDVYRGILVSSGAM 1064
Db 958 KONSATYMKKRHLFMPPIVANIILKRTSNINLIFEVETIRKNVOTIIPDKDNILIPDNV 1017

```

```

QY 1065 VOLAIVAVRAVARGAARQAHMADFMDLSITDDEFLARDLELDDOILLOTLETFWEGALF 1124
Db 1018 IEL-----VKIGSECEMTEEDLOQFYLGYYIMSDEIWSFQULLTDSGTMSVSE---S 1068
QY 1125 AVKILDEKTTACDGGTPTNNLAR 1146
Db 1069 VTKILGSRK-----QEDCNLEP 1085

RESULT 12
S55600
single-stranded DNA binding protein 06 - equine herpesvirus 2
C:Species: equine herpesvirus 2
C:Date: 27-Oct-1995 #sequence revision 03-Nov-1995 #text_change 26 Aug-1999
C:Accession: S55600
R:TelFord, E.A.R.; Watson, M.S.; Aird, H.C.; Peiray, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A:Title: The DNA sequence of equine herpesvirus 2.
A:Reference number: S55594; MUID:95302501; PMID:7783207
A:Accession: S55600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1145 <TEL>
A:Cross-references: GR:020824; NID:q695172; PIDN:AAC13793.1; PID:q695178
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C:Superfamily: herpesvirus DNA-binding protein

Query Match 9.88; Score 618; DB 2; Length 1145;
Best local similarity 23.18; Pred. No. 4e-38;
Matches 295; Conservative 182; Mismatches 540; Indels 260; Caps 47;

QY 4 TOKTYVTGGLGYVYACRVEDLDEELSLAARSTDSDDLPLMRNLYEKTTSLSA 63
Db 17 TOASY-----GPGGVYLYPGDTFPEVEASLIGNHAGEVSLPLLSGLTVEADFNHVK 72
QY 64 VSGARTTGLAGAGITTLKLTSHPEPSVYFVHGKHVLPSSAAPULTACNAREGFES 123
Db 73 AVHKRLDPA-----TVSKASAVHREYIVFANMACFPIFAQVPLBBLVAASRLPQYA 126
QY 124 RCGRPVADGAVETIGA-----ETCRIGLEP--ENTILYVYALFKAVMCMVFLHYG 176
Db 127 EPE-----ERAGGAARPELADLGHLLPGASHLIAGVAVTSEFKERLRCQLVVES 178
QY 177 GLDIIVHNGDVIRIPLFPVOLFMPDVNRKLVPPDPNTIHRSIGEGFYPTFPYNTGCLHL 236
Db 179 QIQSVRNGECDAEKVPRLYDGLFLPAKSPC-----ENLRY---FYHAGVSRV 221
QY 237 IHDCVIAPMAVALRVNVTAVARGAHLAFDENEGAVLPDITYTFYFQSSSSGTTARG 296
Db 222 LFEAHYTSLAQALRVVRVPGILALERQSF---HDYKLP-----KYECREPVATGHRG 273
QY 297 ARRDVNSTSKSPSSGFERLASIMADTALLHAEVIFNTGIEETP--TDIKEMPMFEGM 355
Db 274 A-----GCSLTIIVDSVATELAVSYGLSFLFEPQFQALISYDKMPFIEGC 319
QY 356 EGTPLRNALGSYTAARVAVIGAVFSPNSALYLTVEVDSGMEKDGPGPSFNRFYOF 415
Db 320 ETEPQREVEALTFQNAKQAVHVSOLFSGNSVLYLAPVQKQASN--PAGGGRNNVNSPFG 377
QY 416 AGPRLAANPOTDRG-----HYLSQSTSSNTFSDYALALICGAPUALLRLLFYLER 470
Db 378 HGLACLAEPQKENGLESPFGVPAVALSGSN---YSLHILAVYASFQMLARHCYYLOF 434
QY 471 CDGAFITGGHDAIKYVTGTFDSEIPCSICEKHTRYVCANTYHRLKRMKPR-FGQATRO 529
Db 435 AQHQKSSNNGSVNVPYVIGTAANTPMCELCRGSPPACVNTLFLRLDRPPVVAASVRKD 494
QY 530 PIGVFGIMNSYSDCDPLGNVAYLILRKPGDQTEAKAIMDYATLIERLIDLEDOER 589
Db 495 PYYVTGAGA-YNDLDIAGNFANY---RDKDEESNQEEREKEFTYMOVOTQVL-----ER 545
QY 590 LLDRGACPSSEGLSSVIVDPPTFRRLIDTLARIEQTTQPMKVLVETRDYKIRKGISFA 649

```


Keywords: DNA binding

```

      | | | | |
866 TEPHAILOS MSYNHYSI KOKHAI JYOTLIKPO ATYOK----- 908

```


Cloned from:
Accession: 041
Clustal Family: herpesvirus DNA binding protein

Query Match: 7.8% Score 490.5; DB 2; Length 1132;
Best Local Similarity: 22.8%; Field No. 26 28;
Matches: 278; Conserved: 180; Mismatches: 535; Indels: 273; Gaps: 54;

```

01 2 ENQKIVAVTQDIAVYAVRVEDIMFEISPIAASSTSDALLPMKHEVKEPTSS 61
02 3 DINEIVASAVSTAAIIVYPRKEKELLDVLSVLSIMRNRPVVISLIMLIVENDPSTT 62
03 6.2 LAVVSCAPPTGLAGAVITLKLTTSTPPSVVEFGKINVL-PSSAANPLTRAGNAERF 130
04 6.3 V-----KPTLTFNGGILLIKITSPVCPFFHGTGOLVGMADHDIDLICQHTQK 115
05 1.21 CSPKTCGPVDAVETGACITGICLPENITLIVVIALFEKAVMCMVPLHYGGLD 180
06 116 HLOSFEVPTARKVIDKA--LCSAVGRKDDSVIGIVAGCGKFELEF-----ATL 165
07 181 -----VHINHGIVIRIPVVGLEPMDDVNMIVDPENTHRSIGEGVYPTPYNTG 232
08 166 PVEFQLOVGVGYSVKIPLYSATLLETR-----ETLSQCTHPIQRKCHPLPA 216
09 233 EYHLTHKCVATPMAVAVPRKAVTAVAGCAHLATFENHEFVAVLPPTTYTFQSSSGTT 292
10 217 ESEFLPYVFTSMGTLKPSNKEKIDAGIKPTQDQDQVTKLAPKTYL----- 266
11 293 TACGARRNVNSTSKPSQGFERRIASIMADPTALAEVFN-----TGIVEETPTD 346
12 267 GLSGKISAVEKD-----FLMLVSSV TELSFHVAEYLDVYD--PSQI 309
13 647 -KEMIMETGMGCTIPRINALGSYTARVAGVIGAMVPSNSALVFEV-----D 394
14 310 MNPENMHTLRNSETHAERKAQILNLIKHLSSHIAVLIFAPNSILYCSKLAFIPNVKQAFN 369
15 395 SNTTEAKDKGPPSEKRFYQFAGPULAAVQTDQDRI-VLSQSTGSSNTEFSVDYAL 453
16 370 SVMTQ-----ELDKSLSPFNALSSITFDVYNDRKIKKIDSTSGKDKFSANHLAYA 422
17 454 TCGGALITLALILEYLRKIDAGAFDGGHAKLVYCTPDEI-----PQSLCEKH 503
18 423 CATCPQLTSYVAVNLRKMSV-----YNAGNAVTEIYNHLVNSANLCFECQDK 470
19 504 TRPVTAHTTVHRIKQMPFPGQATROPICVPGIMNSQYSKQDPLGNVATPYLLRKPGQDT 563
20 471 TQVSTTGTAMAVVGIKIDAPLNKKKPIVMSKSPYAAVDIIGSPG-----RKPVS 525
21 564 EAAKATMQDTYATLEKFLIDLEQERLDRGA-----PQSEGEISVIVDHPTR 613
22 526 KETGKQKQNTLS-----LDKGFVSOIPDYCKKNSLIDVTGHDTPNVS 570
23 614 KILDEHAKIHQITQPMKVLVEIRKYI-KKQISFALSMALIDPYSCAPCI 667
24 571 KKDFVSTIHDTQGLEHVS-RCTVENRFPQTPREQIENCLQSFNVDTTPVATFSP 627
25 668 TNPVAKRTHAAVODIALSCCHVFYGGQVEHKNFNOFPIARRKPVLF-----N 719
26 628 LIFATYKVLILVIONAL-----VASGHVVDKPTQNSISKMLVOOYOSLYGTFHSSYLK 683
27 720 GSEFISTRTITVLTSEGVSAAPNPTLQ--DA---PAGR---TFDGLARVSEVIRDIR 770
28 684 KQFLNTRIVK -ASNVDMQIHDGILYKSGKYVKTITQAKICRLSMQCLDRFR 735
29 771 VKNRVVFSQNTINLSAARLVGLASAVQOKRKYMDLHCAICPLIKQFHGILFPRCMP 830
30 736 KNRPNKSSKTAIINPYEKKV-----KHKKNPLSGCISFLFLFYHDKLF----- 781
31 841 RNSK SPINQVFWITLQORQUMADKLTHEETL-AAVKKRFEFVAAINPILNPPCIGR 888
32 782 INVKISLHLMQRPILNNVPTKIIDGNPREVKTIKFASTINTYDEIIDIQRECLST 841
33 889 LAQFYMANLILKYCHDSYL-INTLS-IITG-----ARRRDPSS-SVLIMIRK 934

```

```

DB 842 FIDCYFENKFLSLAGFDYILISLGLSKLVTQNPVLPVAVLLQKPKFSSIGLELYVVK 901
QY 935 DVTSADILEQAKALLKRTENLPLMTTAPTSHLVAAANORPMVVICISIKYHGAAG 994
DB 902 LVLDG--VPNPVYASLSKEBN---FQITFS-----RSLVTFGLLEKFSVLA 944
QY 995 NNHFQAG--NMSGILNS-GKNVCP-----TFDPTFRFTNPRGGFTQPTGSPSSN 1045
DB 945 NREYFOGOLWIGGSGVDKMLNPTSSALQDPFRMR-QKTLIAKPSVIV-----KKVR 998
QY 1046 RETLSQV--VKSIIYS-----GAMVOLAIVATVAVAGAAQIMAFDDMLSLTD-E 1096
DB 999 RETIMEDEVVKKVSLIVENLTNDIDPELLIIAEVVR---DREDKPTMDMLFEVDGRE 1055
QY 1097 FIARD-----LELHDOIIQTLTPWTVGALAV-----KILDKTTAGDG--- 1138
DB 1056 ALAASIMLKLNHLVDMNVRDPSIA-NIQSVFEAVSSNDAPVDFSEILAEEDQASGLK 1114
QY 1139 --ETPT 1142
DB 1115 CDELET 1120

```

Search completed: March 28, 2003, 13:39:02
Job time: 60 secs

QY 592 DNCA-----PCSSGGLSVYVDHPFRRIIDPILRARIDOTTQPMKVLVETRDYKLRIGI 646
 Db 597 DEKALAHAGTGCAS---TGVVADQAFINILSTIKDTTEGAEQFMILVEVDFKIREGL 654
 QY 647 SEATHSMALTDTPYSGACFPITNELVKRTHLAVVDLALSCCHCVYGVQVEGRERNOF 706
 Db 655 ADANHTMSISLDPSYSSSCPVTSFISKRRTIFAVLDLVLSCQCHLFCYGVSGVGRNRNOF 714
 QY 707 QPVLRRRVRVDLFNGSEFISTRITVTLSSECPVSAAPNPLTGQDAPARTPEGDILARVSEVI 766
 Db 715 QPVLRRRRLDMLNGSFTAKTIVTVVSDGVTAAPNLITPSSEPPKIDYGDGMARVSMEL 774
 QY 767 RD1HVKNRVEVSGNCTNI,SEAAARALVGLASAVQROEKRVDM,HCALCPILKOPGLLP 826
 Db 775 RDLRIKNRVEVSGNANSEAAARAVAGASAVRRPEKSNILNCAVGLVQFIKVLFP 834
 QY 827 RGMFNSNSPPOWFTVTLQKQMAFKLTHEITTLAIVKRETEYALINFINLPTCI 886
 Db 835 RGHPRGIDTPNPOWFTVILORNOIPARILSKEDIPITAIKRFSHYSAINFINLPTCI 894
 QY 887 GELAQFYMANLILKXCDHSOYLINTLSTTGARPRDPSSVLTWHIRKQVTSADIETQA 946
 Db 895 GELAQFYANLVLKXCDHSOYFINLSTIIVGSRPRPAVLANIKRTINCAVDEPAA 954
 QY 947 KALLKTEINLPLMTTASTSHILVAAANNORPVLVLSISKYHGAAGNNRVYQAGNNSG 1006
 Db 955 QEVLDQGLSNPAAMVGTFASTINMRYVMDORPMVYIGLISISKYSGAGNNRVYQAGNNSG 1014
 QY 1007 LNGCKNNVPLETFEDTRFETIACPRGFCIPVTSNGNRETTLSDOVIGTIVSGAMVQ 1066
 Db 1015 LNGCKNNVPLETFEDTRFETIACPRGFCIPVTSNGNRETTLSDOVIGTIVSGAMVQ 1074
 QY 1067 LAIYAVYRAVGAQAQIHAFDMLSLTDEFLARDELEHDIQITLEPTVEGALBAV 1126
 Db 1075 TAVESVYVLTALGARTQHLAVDEMLIGLVDEFLAASLDALNAVVOVF--GEMSVGAQAQMI 1133
 QY 1127 KILDEKT-----TAGDGTPTINLAPNPDSCERSHDTJNSVLNISSNSISGTVPLKRP 1181
 Db 1134 RTMAQNTINMGVSTGDS-----AFDGLVGVDAUNSSSTTNMGA--SSSAPAGQKRPH 1185
 QY 1182 EDELFELSLGIPIRKHNITMEM 1203
 Db 1186 PDDILFDMGAPPEKKSGILTPDM 1207
 RESULT 3
 Q89549 PRELIMINARY: PRT: 1203 AA.
 Q89549
 Q89549: 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 GN 01-29.
 OS hOvthe herpesvirus 1.
 OC viruses, dsDNA viruses, no RNA stage; Herpesviridae;
 UC Alphaherpesvirinae; Varicellovirus.
 QX NCBI_TaxID=10320;
 RN [1]
 RP SOURCE: FROM N.A.
 RC STRAIN-COOPER;
 RA Schwyzer M., Vleck C., Lowery D.E., Bellio L.J., Meyer G., Mista V.,
 RL Submitted (MUG-1996) to the EMBL/Genbank/DBD databases.
 RN [2]
 RP SOURCE: FROM N.A.
 RC STRAIN-COOPER;
 RA Meyer G., Vleck C., Lowery D.E., Pastorer P., Thury F., Schwyzer M.,
 RL Submitted (JAN-1996) to the EMBL/Genbank/DBD databases.
 RN [3]
 RP SOURCE: FROM N.A.
 RC STRAIN-JURA;
 RA Schwyzer M., Vleck C., Lowery D.E., Bellio L.J., Meyer G., Mista V.,

RL Submitted (NOV-1997) to the EMBL/Genbank/DBD databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JURA;
 RA Schwyzer M.;
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBD databases.
 DR EMBL: 278205; CA001596.1;
 DR EMBL: X94677; CA064336.1;
 DR EMBL: A004801; CA06104.1;
 DR InterPro: IPR000635; Viral_DNA_bind.
 DR Pfam: PF00747; Viral_DNA_dp.1.
 SQ SEQUENCE 1203 AA; 127409 MW; 8299D64966A9654F CRC64;
 Query Match 53.6%, Score 3771.5, UB 12, Length 1203;
 Best Local Similarity 53.6%; Pred. No. 5,8e-263;
 Matches 652; Conservative 197; Mismatches 340; Indels 27; Gaps 16;
 QY 1 MENTOKTVVTPGPIGVYVACVEDLDLEISPLAARSTDSIALPLMRNLVTEKTPS 60
 Db 1 MDAAKTVALACGPACFYVVCDAARLDLKDIALAARSHSLAVLPLVKGILVETAFAP 60
 QY 61 SLAVVSGARTLLAAGITLKLITSHFYPSVFEVHGKHVLEPSSAAPNLTRACNAAREF 120
 Db 61 NVAVVAGIKTTGLGAGLAKLTPSHYHNVFVPHGGERLRASTADPNLTRCEQARRR 120
 QY 121 GFSRCQGPVDCAVETTCGAFCTRLGLEPENTILVIVVATLKEANVKNVPLHYGLDI 180
 Db 121 GFSARAGAVDVAETTADICTRAVGASPELALVTLATTEAREKTEYVMKNTLHGGAP 180
 QY 181 VHIHQDVIRIPLPVQVLEFMPDVNRLVPPFNTHRSICEGFVYPTPFYNTGICHLIDC 240
 Db 181 VOVGAGFAVRVPLVYQJLYMPVNVNLEPPNARQVAILGEQALAYPPFYNALCELHCV 240
 QY 241 VTAAPVAVLRVNRNAAVARGAHLAFDENHGAVALPDTITTYFQSSSGTTARGAPPN 300
 Db 241 VGPAAVALRVKTVTLAAVAGAHLAFDESHGAVLPPDVCTVF--DOPAGKQGGKRSAG 299
 QY 301 DVNSTKSPSGCFERRLASIAADTALAEVTFNTGIEEPTDIEKPMFICMGRTLP 360
 Db 300 ADPGAAKAALPGVERRLASVMAADTAVISAAMSTSVDEDEVACVDDPMILGAGADAA 359
 QY 361 RLNALGSTARAGVIGAMVESPNSALYLTVEYDSGMTAKGKGPGSPNRRYOFAGPHL 420
 Db 360 KIDALCAVYGRLAGVAGAVFSSNVLMHTEYDDGGAADAKGA--AAGHRRFYQIAAPYA 418
 QY 421 AANPOTDRDGHVLSQSTG-----SSNTEFSVDYALICGGAFLRLARLTYLERCDG 474
 Db 419 AGNPRGDKDKRLPOTGAPRAVSIAGAGGERFALDIALACGCPOLLAMLYLERCDG 478
 QY 475 AFTGGHG--DAIKYVITGTFDSEITPGLCEKHTRPVCAHTTVHRLRORMPRGQATROPICV 533
 Db 479 AFAGRNDDALKYVASTLEGDVPGLGSRDNRHACHTLHLRLPRLPFGAPTSPLGV 538
 QY 534 FCTMNSOYSDCDPLGNVAPYULIRKQDOTTEAKKATMODTYATLEIRLITDIEKRLDR 593
 Db 539 FCTMNSAATSDCDVIGNVASYSLRPPG--ADEANRSIMQATYRAAVENYALIDQOQILAH 597
 QY 594 GAPCSSEGLSVYVDHPFRRIIDTLRKRIEOTTTOFMKVLVETRDYKIREGLSEATHSM 653
 Db 598 DAOSAAQ--LERAITDHAISFRGLAALIMONTVEQATEAFVGLVEDDFKMRALVYANHTL 656
 QY 654 ALTEPYSAGCPLTNFLVKRTHLAVVDLALSCCHCVYGVQVEGRKKNQFOVLRRR 713
 Db 657 SLADDPYTAACPATASFRRKSVLAVVDLALSCCHGIFCGQVGRKNEARAFQVPLRRR 716
 QY 714 FVDLFGNGFISTRITVTLSSECPVSAAPNPLTGQDAPARTPEGDILARVSEVI 773
 Db 717 FMDLNGSFLITRTVYTLAEAAVAPNLAAGVLRPAKUMIGUJSKSLIEYFKMKRKN 776
 QY 774 RVYFSGNCTINSEARALVGLASAVQROEK--RVDMLHGALGFLDKQGHGLFPGGMPN 832
 Db 777 RVMSAGNANMSEARARAVLGLAGAYOKPESGCVNLTSGPIGLVKGPHRKLIFPMCKPDC 836

069101
ID 069101 PRELIMINARY: PRT: 1197 AA
AC 069101:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE DNA binding protein ICP8.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCH1_Taxid=10310;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=KN:
RX MEDLIN=93228441; PubMed=8385914;
KA Tsch Y., Tanaka S., Liu Y., Mori R.;
RT "Nucleotide sequence of the major DNA binding protein of herpes
RT simplex virus type 2 and comparison with the type 1 counterpart.";
BL Arch. Virol. 129:183-196(1993)
EMBL: D10658; BAA01507.1;
InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_dp.1.
SO SOURCE: 1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;

Query Match 47.9%; Score 3012; DB 12; Length 1197.
Best Local Similarity 49.0%; Pfam No. 6.6e-214;
Matches 595; Conservative 212; Mismatches 368; Indels 40; Gaps 16;

QY 4 TQKTVTVTTCIGYVY--ACRVEDLDLRETSPLAARSTSDIALPLMNTIVKRTTSS 61
DB 7 TTTVKKVPGPGVYVGRACPAEGLEL--LSLSARSGADAVAVPLIGLVESSEFAN 64
QY 62 LAVSGARTGAGAC:TKLKTSHYTSVFVHGCKHVLSSAANLIRACMAANRRC 121
DB 65 VAAVVGSGCTGIGAVSLKIMPSHSVSVYVFGCRHLAPSTQAINLRLCKRARRHC 124
QY 122 FSRGCGPVDCAVETTGAEICTRLDEPENTILYVLTALFEKAVFMCNVELHYSGLDIV 181
DB 125 FSDVAPRCDLKHETTGDAICERLGLDPRALLIVITGCFREAVCSNIFLHICMDKV 184
QY 182 HINHCIVIRILPFPVQIFMDVGNRIVPDPFNHRSIGCFYVPPRYNTGCHLIDCV 241
DB 185 TIGDEVHIRIPLYPLQMTPEDSRVIADEPCNHRSTGENENTPLDEFBRPLARLFEAV 244
QY 242 IAPMAVALRVNVTAVARCAHILAFDENEGAVLPDIIYTYFOSSSGCTTAKARRND 301
DB 245 VGIAAVVALKARNVDAAKAAHILAFDENEGAVLADITFAPEASQC--KQKQK 299
QY 302 VNSTRKPSDSGFERRLASTMAADTALIAEVFNFTGTYEETPTDIKEMPTFGMECTLP 361
DB 300 --DAGNKGAGDGFEOHILASVMAGDIALALEISVMAVFPDEPPDITTWILKGOETPAK 357
QY 362 TNAIGSYTARVGVIGAMVFPNSALYDFEVDISGTFKDKDGPSPFRFPYOFACPHIA 421
DB 368 AGAVGATYLARAGLVGAMVFTNSALHILEVDAGAPADKDKSK--PSTFPFLVPGTHVA 416
QY 422 ANIQDRDCHVL-----SSQSTSSSTNTPSVVYLALICGFOAPILANILEYLFKCDAG 474
DB 417 ANIQDRDCHVAVGVEGPIAPLVGCTQ--EFAGEHLAMICGSPALLAKMLFLERCDG 475
QY 475 AFTGRHGS--DALEKYVTFSTSETPTSLAEKTKPVCAHTIVHLRQMPFEGVATKPLV 533
DB 476 VLVGQEHMVFVYVAVNSCTIIVNCNCFETTHAACAHITLMIRAHHPKFAASAARCAIGV 535
QY 534 FQIMNSQYSLCHPAGNAYVLLIKKQNGFPAAKATMDQVYFATLPRDPLDIFPFLDIP 593
DB 536 FQIMNSQYSLCHPAGNAYVLLIKKQNGFPAAKATMDQVYFATLPRDPLDIFPFLDIP 594
QY 594 GAVPSSGGLSSVIVHPIVFKILLIIPAAVFGTTIGPMKAVLFTVPAKILKQI--STAHSM 653
DB 595 AVPTAIGRLFTIIGKREALHIVVNNIKQIVDRVPMQIMRNLLIEGRNPKFRDGI--AFANHAM 654
QY 654 ALTEFPYSAPFPIINLVKRLHLAVVVDLALISQCHICVYGGVVEGRNFRNCPVLRRR 713

655 SLSPDYTTGAPPLDLLLRKSNLAVYGLALASG--HSAVTAQSVESKRNPNFQPLRRP 714
QY 714 FVDLFGNGPITSPRSTVTVLSEG--PVSAIPNPJLQJAPACHTPIGDIARVSEVYIKDIK 772
DB 715 VMDLFNNGGLSLAKTTLVLAEGNAICAPSLTAGOTPAEESSEFGDVARVLTGPEKRLK 774
QY 773 NPVFSGNCTNLSEARAPLVGLASVYQDERPVIMLHA--SLKLVGFHLLFPPKMPPN 832
DB 775 SRLVFAASANASEAKARAVASLOSAYQKPKDKVDILGLPGLIKQFAVIVPENGKPPG 834
QY 833 SKSPNPEWMTLLQKRNMPADKLHEEITITIAVKKRETEEVAALINFPETICIGELAF 892
DB 835 SNPNPQWMTLQKRNQULPARLSKEDIETIAIKKESLDYGAINFTINLAPNVSELAMY 894
QY 893 YMANLILKYCDHSQYLINTLSITTCARRPDPSSVLMWIKKVTSADIETQAKALIEK 952
DB 895 YMANULIRV--DHSYFINTLITAVIAGSPRPVJAAAMAFQ--SAGLEAGAPALMDS 951
QY 953 TELPELMTATFTSTHUVANAMQREPMVYGISISYHAAAGNNRVEFVAGNMGSLGKN 1012
DB 952 LDHFGAMTSMRASCULRPVMAARPMVYIGLISSTYTGAGDQVFOAGNMASILGKN 1011
QY 1013 VCPPLFTDRTPTIACPPRGFTPYTGPSS--NPFITLSTVPJIVSYAMVJLAIVAT 1072
DB 1012 ACPLLIFFDTRKRVLCAPRAGFCAASLGGAGHESHLCEQLGLIIAEGGAIVASVEVA 1071
QY 1073 VVAAGVARQAHAFDWMLSITDEFLARDLEIHDQIQTLEFPVIVEGL-----FAVKI 1128
DB 1072 TKSLSGPRQVULIEDMLLLEDEYLSSEMEFTTIALEGHGEMSTDALAEVAHDEAL 1131
QY 1129 LDEKTAGDFTPTNTAFNEDSCPEPSIDTTSNLTNSGNSGSTVPGKRPPEDELF 1188
DB 1132 VSLD--SAGH-----VFNNGVHGHGHHHAASFGILAA--AAAGAVARHPARHDDPFC 1183
QY 1189 LSGIPKIGNTMEM 1203
DB 1184 -ESPEPK--TITLDM 1196

RESULT 6
ID Q9P252 PRELIMINARY: PRT: 1190 AA.
AC Q9P252:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE Homolog of HSV-1 glycoprotein.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCH1_Taxid=10390;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=HPRS24;
KA Kato K., Jiang H., Izumiya Y., Cai J., Tsushima Y., Miyazawa T.,
RT "Identification and Transcriptional Analysis of the Marek's Disease
RT Virus Simplex 2 Genes Homologous to the Glycoprotein B (UL27), the
RT ICP18.5 (UL28) and the Major DNA-binding Protein (UL29) Genes of
RT Herpes Simplex Virus Type 1.";
FL J. Vet. Med. Sci. 0:0-0(1999).
EMBL: AB024711; BAA83753.1;
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_dp.1.
SO SEQUENCE: 1190 AA; 130339 MW; D7291E3EAA7F5AD CRC64;

Query Match 44.4%; Score 2796.5; DB 12; Length 1190;
Best Local Similarity 45.3%; Pfam No. 1.7e-216;
Matches 546; Conservative 245; Mismatches 396; Indels 29; Gaps 13;

QY 6 KTVVTPGPGVYACRVEDLDEELISPLAARSTDDLALPLMRLNTVKTTSIAVY 65


```

QY 306 SKPSPSGEERFLASTMAADTLHAIEVINTGIEETPIDIKEMPMFIGESTLPRNL 365
DB 304 S-----GGTERRKRLASIMASDAILSTINAVIATSVESIPDKWPKPLCSIGTTDVEAL 359
QY 366 GSYTARAVAGVIGAMVSPSALYLIEVEDSGMTEAKD-GGPPSPFNKFFVAFAPHLAANP 424
DB 360 SAYMARVAGLVGAMVSSSVIYMTEVGESSADGDTSTTASFRFQIAPHLASNP 419
QY 425 QTPRHGHVLS-----SOSTSSNTHSVLYALICGCGAPILARLLPYLRGACATGCH 480
DB 420 LTRKGRKPIRGEDLSKATISASPEVSLDYLILACGPPULAPLRFYLEREDGSAACHH 479
QY 481 G-DALKYVGTGDSSEIPECSICEKHTPRVCAHTVHRLRQMPREGGATROPICVFGMNS 539
DB 480 DLDYKRVSSAMDADVPCHLCKASKRIYCAHTTKRLERHLPFGYOMKAMGLPSMTN 539
QY 540 QYSLCDPLGNVAPYLLLRKPDQTEAAKATMODIYRATLERLIDLEUPRLDKAPSS 544
DB 540 NYCDVVALGSYAPSTLKR--SEGFTSRVMDTYKLTVERVAKALEKGLTCEDDPADM 597
QY 600 KGLSSVIVHPHFRRLIDTLRAKIHOTTFQFMKVLVETRYKIRKGLSFAHSMALTFDP 659
DB 598 TPADALTRDGKSFMRALSTMANIIESEAGQLMKNTLREYNIKEGLGADHILSLAIEP 657
QY 660 YSGAEPIPINFLVKRTHLAVVLDLALSCVHCVYGVGVSEGRNRFQVPLRRFVDFLN 719
DB 658 YSSGICVPLSFLSKRTIIVAVQDMALSGCSMIMHQOVERNRTQPOVALRRRVLDION 717
QY 720 GGFSTRITVTLSEGVVSAPELTLQDAPAGKTFEGDLARVSEVYRIDIKVKNVESG 779
DB 718 AGFTISKNTVLEDDQIISVPDFSKQHDPLAVHMEGDLVKTFFELFKREKVKKNKVFVG 777
QY 780 NCICN-LSFAARALVGLASAYQOKERKRVMLHGLGFLLKQFGLLPPRRMPNSNSPNP 838
DB 778 GVSSTVSDATKSLAGMIEAYQPPAKAMHVLNGLFALKRHYTLQLEPVNKKMGNGTTPNA 837
QY 839 GWFETLLQNPADKLTHEEITTAIVKRFTEEYAINFINLPTCIGELAGFYMANLI 898
DB 838 LMFVILLQNOILPACILISDAEENTSFIRKPTNSYADMTINISPTCFGLAQFYLANLI 897
QY 899 LAYTTHSOYLINTLTSTTGAPRPPSSVLMIRKDVTSADITQAKALLEKTENLPE 958
DB 898 LKYSCKHKEFETISALVAVISKPRDPAVLPWIEHILTOGSDVAAVQDLLENIDHKE 957
QY 959 LMTTATSTHILRAAMNOKRMVVLGISIKYHGAAGNNKRFVAGNMSGLNVCNCLFI 1018
DB 958 AMCAAFSSNINLGFIMASKPEVYIVANSISKYHGMASSTKVFUSGMNINIMGRNVSIMS 1017
QY 1019 FDRTRFTIACPPGAFICPVYTGSSGNPFETLSDOVRGIIIVSGAMVQIATATVVPVAV 1078
DB 1018 FMTTRFVACPRVGVSHQTFSSGCKETTLVDRARAVISECGAPHAAYVMAIALKMGV 1077
QY 1079 ARQVIMAFDMLSLTUDELAKLELHDQIITLETPTWVEGALAEVAKLLEKTTAGDG 1138
DB 1078 DRYRQELDDMMIINDAVISSLIDELNKV-VEGCEGSGVSDAAAILLAKMVMMSKRLPL 1136
QY 1139 EYPTINIAFNDUSCEPSHOUTSNVILNSGMSISSTVPGKRPPEDEDLFPLSGIPIKHG 1197
DB 1137 DGPT---FNYDALDENTERRAD-----GFSI---LEPTLKRPSD--VFOLFVPERKAP 1183
QY 1198 NITMEM 1203
DB 1184 GLSVDM 1189

```

```

RESULT 8
Q96P0 PRELIMINARY: PRT: 1191 AA
AC Q96P0:
DT 01-MAR-2001 (TREMPLrel, 16, Created)
DT 01-MAR-2001 (TREMPLrel, 16, Last sequence update)
DT 01-JUN-2001 (TREMPLrel, 17, Last annotation update)

```

```

DE U129 single stranded DNA binding protein-like protein.
GN MDV042.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=103390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD5;
RX MEDLINE=20392152; PubMed=10933706;
RA Tulman E.R., Alfonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
PT "The genome of a very virulent Marek's disease virus."
RJ J. Virol. 74:7980-7988(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MD5;
RA Tulman E.R., Alfonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases
DR EMBL: AE243438; AMG1422.1;
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_bp.1;
SQ SHQUNCE 1191 AA; 130714 MW; 9AD425780F807711 CXC64;

Query Match 44.3%; Score 2788; DB 12; Length 1191;
Best Local Similarity 45.1%; Pred. No. 8,4e-216;
Matches 542; Conservative 237; Mismatches 395; Indels 28; Gaps 12;

QY 1 MENTOKTVYPTPGLVVACREDDLDLEISFLAANSTIDIALPIMKNTLYKRTFS 60
DB 1 MDGVGSSVKLCGGPIGIVYIPKCSVPEDELALMAKSNCCDVAFLVAGLVESDFW 60
QY 61 SLAVVSCARTTGTAGAGITKLTSHFYPSEVFEHGKHLVPSAANLRACNAAREF 120
DB 61 NVAAVACTKTTGCGSGTTIKLPVTHPCVFPYGGDCIKPCKAINTKACDIARERF 120
QY 121 GYS-AVSSPAFTFEFTGGQICELEMDAONVMYLIVTELFKEVILYCNSELFHGS 179
DB 121 GYS-AVSSPAFTFEFTGGQICELEMDAONVMYLIVTELFKEVILYCNSELFHGS 179
QY 181 VHNHGVILKLPFVPVGLHPDVNKLVPDPNTHHRSIGCFVYPTTFYNTGCHLHIC 240
DB 180 VTNINNAVPRIPYPLHLVLPDPNPITNEPSEKFRALGEGALIPKAFNDSTRLHGY 239
QY 241 VIAPMAVLRVAVNTAVARAAMHLAFDENHEGVALPDITVYTGSSSSSTTAFAPN 309
DB 240 VLSITVAGLKVKNILDAIKAAHLCHPBNHSTLLPADITFIATTAAR---TKKQSKM 296
QY 301 DVNSTKPSPSAIFERRLASIMAACTLAHAEVINTGIEETPIDIKEMPMFIGESTL 360
DB 297 GKREGSDVS-GGGEYERTASIMASDAILSTINAVIATSVESIPDKWPKPLCSIG 355
QY 361 RLNAIGSYTARVAGVIGAMVSPSALYLIEVEDSGMTEAKD-GGPPSPFNKFFVAF 419
DB 356 KLELSATMARVAGLVGAMVSSSVIYMTEVGAGSTEGEISTVAPSTFRFQIAPHL 415
QY 420 LAANPQTRDGHVLS-----SOSTSSNTHSVLYALICGCGAPILARLLPYLERC 475
DB 416 LSNAPLVDRKGRKPVSGNLSKISASQSEVSLDYLILACGCPQILARLLPYLERC 475
QY 476 FTGCHG-DALKYVGTGDSSEIPEVSEKHTPRVCAHTVHRLRQMPREGGATROP 534
DB 476 KACHHDLDTYKRVSSALIDADMPCELCDKTSRIYCAHTTKRLERHLPFGYOM 535
QY 535 GTMNSGVSTDPDPLGNVAPYLLLRKPDQTEAAKATMODIYRATLERLIDLEUPRL 594
DB 536 GAMTNATCDVVALGSYAPSTLKR--SEGFTSRVMDTYKLTVERVAKALEKGLT 593
QY 595 APCSSSEGLSSVIVDHPFRRLIDTLRAKIHOTTFQFMKVLVETRYKIRKGLSFA 654
DB 594 DPTNMAADANIRHDSNFIATSTKRNISEASQLMKNTLREYNIKEGLGADHILSL 653
QY 655 LTRPPYSTARVPTNPLVKTHLAVVLDLALSCVHCVYGVGVSEGRNRFQVPLRRF 714

```

[illegible]

QY	1	MLNTOKTYVTTPGPIGLGYACRVEDJDLHETSLAASTSDJALLPLMKNLTVEKTPUS	60
Dh	1	MGVGKSVKLVGGPFGVYYSKSKAVVEDELSTFMAAGJISESLPLPLGLTVESEFVR	60
QY	61	SLAUVSGARTTGLGAGITLKLITSHPEYSEVFEHGKIVLPSSAAPNLTRCNAREVE	120
Dh	61	NVAAGAATIKTIGQVIMGOTITIKLVPIHHNVPVYGGGLNCSKAPNLNACIYAREEP	120
QY	121	GESKQGPVPDGAFTTGAETCTPLGLEPENTILYLVYALFEAVFMNVFLHYGGELDI	180
Dh	121	GYCPYSAP---GSVEISKQICEDLGIDACKTMYLYVLETFEALYLCNSFYGGSDS	177
QY	181	VHINNGDVILPELPEVOLFMPDMVNRILPDPINPHRSIGCFVYPIPHYNTQICHLHQC	240
Dh	178	VTINDAEVRRIPLPLHLYLVEDFKRVSNNEPSEPRALGGAAVPEAFNDALCRLLYAY	237
QY	241	VIAPAAVALRVBNVTAARGAAHLAFEDNHEGAVLPDITITYF---QSSSSSTTTARGAR	298
Dh	238	VLGTDAVLKLRNIDVAARGAAHLCPENHGHLLLPADPTFLAPTPADTGKQCKACK	297
QY	299	RNDVNSTKSPSPSGGFEERLSTAAATLHALEYFNTGTYEETPLDIKEMPEFTIMEGT	358
Dh	298	RDGAEE-----TGGGVERRTASLMASTATLAEVLAASSVEEPIDPVKKWPIFCNPVAY	352
QY	359	LPRNALGCTYARAGVYGAWVSPNSALILFEVNSGMEFAKDQC-PCHSENFVQFAG	417
Dh	353	IDREVALSAYMGVAGLVGAWVFSSNVITYETVEGASSDCKGSEGVAPASFEFRFQJAA	412
QY	418	PILLANPOTDEPDRHIVLSNQS---TSSNTEFSTVYALITGGAGELAEFLYLEKCA	473
Dh	413	PRILANPLVYDDKGRILSDALAPKLAVSTSSYSLDLYILAGCQVQILAAFLPYLPCRCQ	472
QY	474	GAFTEGNGDALRYVTTFDESEIPGSECKEHTPRVANHIVHLPQMPPEFATPQPG	522
Dh	473	GANNCHHDLDVTKLVSSAIDADMPCELDKTSRYVCAHTTKRELEVRLEKPRVQMGHAG	532
QY	533	VWGTNNSQYSGNDPLGNAPRLIRKQDQTEAAKATMONTATLPRLEFIDLEQRLD	592
Dh	533	LEGAATNNQCVNALGSAQSTLKR---SGEASRSVMODIYPLIYERMAKLEKGLLA	590
QY	593	RGAPSSSEGLSSVIVDPHTPRILDTLPARIEATTPFMVIVETPRVYKIPELSEATIS	652
Dh	591	CDHPNNMAPALASIKDQSPFHAASTMRNILEGASQJLMMNLTEIKRYNKRQICIANHF	650
QY	653	MAITFDPSGAFCTPTNTFYKRTHLAAVYDLASQHFVYQVGFENFPMQFPLVPIPR	712
Dh	651	LSLTIEPSSSPPEVLSFLAKRTIAVLDQMALSYQSMLEHSGVFAVNEPRTAFQSVLEPR	710
QY	713	RPVDFLFCNGFISTFSTIVILISGCVSAPNPLDQDALACHFICQDLARVSVFIRQIYVK	772
Dh	711	RYLELQANGFTTSKDTITLFDQIVAVAPDPSRSQYDPTVYINMGDLTRVITIELIRELKK	770
QY	773	NRVYESGNCT-NLSEAPARIVGTASAYQFQEKRVUMLHGLALGFLLEKPHGLFFRGMPR	831
Dh	771	NKVLIFQGGVIGISPADAKSKILANNIEAYQPRKTPMNVINCPGFAIKRVHSAFLPNVMPK	830
QY	832	NKSPNQWEMETLLQBRMMPADKLTHEITITIAVVRFEFEVATINFTNLPTQIGELAQ	891
Dh	831	NGAVPNAHMEFTLLQKRNULPOTKLSKELCEKALFIKFTNVADMMVYNISPLGRDLAQ	890
QY	892	FYMANLILIKYDHSQVYLNTLTSILIGARRPRDPSSVLAHIRKDIVISADVLEQKALKLE	951
Dh	891	FELANTILIKYCSIKHFFINTISALVAASKPRDPAIVLPIKTIQVGRDVASLMOGLIG	950
QY	952	KTENLEPLMTATLFTSLVLRANNOBPVYVLGISIKYHGAAGNNRVDFVAGMSSLNGKR	1011
Dh	951	TIAJRKQAWCATPFSATINIVGSVSTKPIIVIGVSISKYHOMASCTKVPSQGMCMIGCR	1010
QY	1012	NVCPLETFEDRIRREILICPGGGEICPEYTGSSSNRETTISDQVRCITVSGGAMVJLATA	1071
Dh	1011	NVCSLMSEDRHRRYVACPRVGEVALEITFSSSLKEKENTLMDAKAILSEBAGAAAEVYM	1070
QY	1072	TVVRAVGAARQAHMAFDIMLSITJDEHFIARDILEHDOILQITLFTWTVGALFAVKITIDE	1131

Marek disease virus encodes a basic leucine zipper gene resembling the fos/jun oncogenes that is highly expressed in lymphoblastoid tumors.

RT. Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).

RM [2]

RD SEQUENCE FROM N.A.

RE STRAIN CA:

RA MEDLINE 20283955; PubMed-10824954;

RB Lee L.F., Wu P., Su D., Ren D., Kamil J., King H.T., Witter R.L.,

RT The complete unique long sequence and the complete genomic

RE organization of the CA strain of Marek's disease virus.*

RM Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096(2000).

RD PubMed: 1191806; AAF6764.1;

RE InterPro: IPR000645; Viral_DNA_bind.

RM Pfam: PF00747; Viral_DNA_bp.1;

RD SEQUENCE: 1191 AA; 140946 MW; AC04A015B0797A3 CRC64;

SO

Query Match 44.1%; Score 2710; DB 12; Length 1191;

Best Local Similarity 44.1%; Pred. No. 1.7e-209;

Motif 545; Conservative 246; Mismatches 400; Indels 42; Gaps 16;

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

1 MENTKIVTPTPTPLVYACVVEDELEISE-----LAARSTSDALLPLMNTLV 54

768 DLKKNKVVVSGNCLNLSAARAKVLGLASAVQROKRVDMHGAFLKQFHLFP 826

767 ELKKNKVVLEGGITACASASATSRILAGVAVQPTKTHVNLGDLGFAVKKRYHLFP 826

827 KGMPPKSPNPQWFTLLQPNMPADKLTHEETITAAVPEETEEVAAINLPPLT 886

827 DVMKPGATPINALMFWLILRLNOLPACILSKFEEDSLFLEKPKTSADNVAINISPT 886

887 GELAGVYMNILIKYVTHSLYCLNTLSITVAAPPPPSVLMIPKOVTSAADETQA 946

887 GDLQVYLLNTILIKYSIHKEPINTLSALVASSRRPPDAIVLWLEPRTKGDVAPA 946

947 KALIEKTEMLPELMTAPSTHILVRAAMNORPVAVIGISIKHYCAAGNRRVQACNMSC 1006

947 QDLIASMSHKDLMCATFSSINLVGIMTKPFVYIGISIKHYCAAGNRRVQACNMSC 1006

1007 LNSGKVCPLFTFDKTRKELIACPRGGLCPVTGPSSGNRETLISQVRCIIVSGAMVO 1066

1007 IMCGNRVCSIMSDRTHRYVMTCPRVGFAVQOPFSSGKIKETLLIDRVHVLSESSAAR 1066

1067 LAIVYAVAVGARAQHMADFDMLSLIDDEFLARDELDHDIOTLEPMTVEGLEAV 1126

1067 AAVYMLALRMVGDVROMELDMETINDEYISLIDELKQVEEABGGMNADAMTLA 1125

1127 KLIDEKTAGDCTPYNLAAPNPDSCPSHDPTNSVNLISGNSISGTVGLAKRPDEHL 1186

1126 KENVNMAKSIPTDGP---FDFACDENLEGHADGQISBTN-----LKRPNMNV 1172

1187 FDLGSPKIKHGN 1199

1173 FDLPIPKRAV 1185

RESULT 12

090H63

1D 090H63 PHELMIMANKY PRT 999 AA.

AC 090H63

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 17, Last annotation update)

DE DNA binding protein.

OS Gallid herpesvirus 1.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.

OX NCBI_TaxID=10386;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SA-2;

RA Johnson M.A.;

RT *Sequence of the infectious laryngotracheitis virus (SA-2 strain)

RT unique long region DL28 to DL41.*

RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF168792; AAD56201.1;

DR InterPro: IPR000635; Viral_DNA_bind.

DR Pfam: PF00747; Viral_DNA_bp.1;

SO SEQUENCE 999 AA; 108477 MW; 089FADP7C8H674H CRC64;

Query Match 23.0%; Score 1445; DB 12; Length 999;

Best Local Similarity 30.5%; Pred. No. 2.3e-107;

Matches 351; Conservative 177; Mismatches 376; Indels 246; Gaps 18;

13 GGLGVYVACRVDDIDLEISPLAASSTDS-DIALPLMNTLVEKPTTSSIAVSGARTT 71

22 GFLGVYVAFREEDIDGKEKILCAKSQDPSCAIAPLILGLTVESEDFKFNVAVITKSS 81

72 GLACAGITLKLITSHYVPSVVFHCGKHVLPSSAANPLFRACNAARERFGSGC-GCPV 130

82 GVGCGNCAIALISPCHPSTTVTFHCGKCPITTSFTPLNLKICHDAKRRKFGSSIPNCGV 141

131 DGAVETTAGCITRLGLEPENTLVLVYALPKFAVFMKCNVFLHGGDILVHINHDVTR 190

142 PNALETGHEICKSINMDPKIMLVYVAHPCEAVAVCNVLIHGSVYLIVINSKHTK 201

```

QY 191 TPLFVQLFMPDVN-RLVDPDPNTHHRSIGRCVYVPTPEYNTGLCHLJHDCVJAPMAVAL 249
Db 202 IPIYVQVMMDDIARICRNPFDNSKINGESCITPTPLFNKALRKLHGAILTGGSGL 261
QY 250 RVRNVTAVARCAHILAFDENEGAVLPPDIYTYFQSSSSGTTTARGARRDNVNSTKPS 309
Db 262 RTKDLFAVARCAAMLAIFDSCFECVLTSDKFTTEL 307
QY 310 PSGFERBLATIMADTALHAEVINTGTYETPRD-----IKEMPMFISGEGTLPRLN 363
Db 308 QNVEYERAAASLAELALATRLYSAPY---PBDNGCGDLYEEMPLECDCKTOMERO 364
QY 364 ALGSYTAHVAVICAMVHSPNSALYLTVEVDSGMTAAKDGPG---PSFNAPYOPAGH 419
Db 365 ALDFRSALAGIVGAALFISNNSPLVASEVDGAP-----PGEASSGLTFEYLTGGLH 419
QY 420 LAAMPVJDRDH-----VLSSUSTGSSNTEFSVDYLLALICGGAFLARLLEYLRCGAG 474
Db 420 LIGCGVAVYSGKRWEGTGPSPILTPQNTFFHYDPEHAAVACGSPHLSNIIFFYLEKSKO 479
QY 475 AFTGGHGDALKYVTGTETSEIPCS-----LCEKHTPRVCANTHVRHROBMPREGAT 527
Db 480 LSTRN---MMLISTNSGKAPCTGAVPECKMGNDESNNHCTRTYMOHLPTFTFPIPTPR 536
QY 528 KQPIGVCTMNSQVSDCPDIGNVARYLLLRKPDQTEAKAKATMODYRATLEERLEIDLE 587
Db 537 RSPAVIAVAGVGDYTDCCILGNFAVYSQLKTRAGD-GEPAKAAVNDYTRKIGRWVFOYLS 595
QY 588 ERIDRGAFCPSSEGLSVTVHPTFRPLDILRAIEDTTOFMKVLVETROYKIREGLS 647
Db 596 EGLINKDIDGETHNSL-----AEIKTYFEKIONFVGESECSFISALSGVARYHYKHLIS 651
QY 648 EATHSMALTEDPYSGAFCTPNTFLVTRTHLAVYDOLASQCHVYGGQVEGRNFRNOFO 707
Db 652 SIATFG-SPEPYSSAYCPVLSLLCAQTRSLIFODLILISQIHGTDTPTOPESKMRSSAM 710
QY 708 PVLRKKRVUJLNGCPILSTRITVILSRGPVAPNPTLQODAPAGTTPGDLARVAVEVIR 767
Db 711 PALSAENGMIDKQFLSGKYEPFISVPCVNAPTMRPTEOALAQVYSLSRGVLELK 770
QY 768 CIPVKNPVFVSNTNLSEAAARLVGLASAYOROEKRVMDLHGALGFLKQFHGLLEFR 827
Db 771 EFKVKNRVFVNGSSSNS----- 788
QY 828 GMPNSKSPNQWETLLQKRMADKLTHEITIAVKKFTEYAAINFINDPTCIG 887
Db 789 ----- 788
QY 888 ELAOFYMANLILKYCDHSOYLINTLTSITGARRPRDPSSVILWIRKDVTSADILETOAK 947
Db 789 ----- 788
QY 948 ALLKTEKTELPMLTATFTSHLVRAMNORPMVVLGISTKYHGAGANNRVEQACNWSGL 1007
Db 789 ----- 789
QY 1008 ---NGKNVCPLEFTRRTFTLTPRGGFTCPVTPGSSGKRETTLSUVGILLVSGAM 1064
Db 827 IGNSSQSVNSALINDYPRKALACKRVGILLSSQVAVSGOEINTASOVNSLIDAGS- 885
QY 1065 VOLATVAVARAGARQOHMAFDMLSLTDEFLARLDLEIHDQIIGTLETFMVEGAL 1124
Db 886 -PTSIYTAVALRVLGEKMDIAETMAITDCKYLVNLVELREDISGS-QNGMSVATIOS 943
QY 1125 AVKILDEKIT 1134
Db 944 MLSRLBCKDT 953

```

RESULT 13
ID 040913 PRHLMINARY: PRT: 1132 AA.
AC 040913:

```

DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE ORF 06.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
RT Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
RT human herpesvirus 8: determinants of its pathogenicity?";
RT J. Virol. 71:4187-4192(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Neipel F., Albrecht J.C., Fauset A., Huang Y.-C., Li J.-J.,
RA Friedmann-Kien A.B., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
RT Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, U93872; AB62603.1; -;
DR InterPro: IPR002114; HPr_Serp_site.
DR InterPro: IPR001993; Mitoch_carrier.
DR InterPro: IPR000635; Mitoch_carrier.
DR Pfam: PF00747; Viral_DNA_bind.
DR PROSITE: PS00215; MITOCH_CARRIER. UNKNOWN_1.
DR PROSITE: PS00589; PTS_HPR_SER. UNKNOWN_1.
SO SEQUENCE 1132 AA; 125394 MW; CF88E4F451919CBF CRC64;

Query Match 10.9%; Score 683.5; DA 12; Length 1132;
Best Local Similarity 23.5%; Pred. NO. 1e-45;
Matches 301; Conservative 196; Mismatches 500; Indels 265; Gaps 53;

QY 11 PEGPLGYVACVEDLDELISFLARSDSLPLMKNTLVKPTFTSSIAVVSAGAT 70
Db 18 PEGPGVLYAYLILHNPPIGEASILNGVPEAKVSLPLDGLTVDSPFLNVKAVH--- 73
QY 71 TGLAAGITLKLKLTSHFVSFVFGHGKHLVLPSSAANPLTRCANAREFGFSRGGPPV 130
Db 74 KRIDATTSVAKLTSYH--REALVFINTHLFQPIFGKGLKLCRESRELFGSTVEEOH 131
QY 131 DCAVETTGAEICTRIGLEPENTILYLVATLKEAVFMGNVFLHYGGLD)-----VH 182
Db 132 KQTL--WSPEDACPOLPCANE-IFMAVITYEGFKERL-----YGGKLVVPQSQTPIVH 180
QY 183 INHGDVIRLPEPVQOLFMPDVNRLVDPDPNTHHRSIGSGFYVPTPEYNTGLCHLJHDCV 242
Db 181 IGEHAPFKIPLADEDLFGPSRAQEL-----CRPYNDIISRYLHDSITP 222
QY 243 APMVAALRVNRYTAVARCAHILAFDENEGAVLPPDIYTYF-QSSSSGTTTARGAFPMN 301
Db 223 TGLAQLRLKVESTVIO--ASERQF--VHDYKIFPLVYAKQDFQCAHSRITDQ----- 271
QY 302 VNSTSKPSPSGCFERKLASIMADTALHAEVIFNNGI-VEETPTDIKE-----WPMPTIG 354
Db 272 -----STLWVID-SLVAELGMSGLSTESIEGPODSCVELVNDYTPWFEN 313
QY 355 MEGTLPRNLALGSYTAHVAVAGVAVFSPNSALYLTVEVDSGMTAAKDGPGPSNFRFQ 414
Db 314 CETPRARLRALEVMHARQALHIGALFLAANSVLYLTRV--ALPQKORGDAMNMSNFTL 371
QY 415 FAGPHLANPOTDRG-----HVLSSQSTGSSNTEFSVDYLLALICGGAFLARLLEYL 469
Db 372 QHGLGYLSEATVKEGASAFKGVPSALDSS--YTLOHLAYVASSFPHLLARMKYU 428
QY 470 RCDACAPFGCHG-DALKYVTGTFPDESEI PCSICEKHTPRVCATTHVRIKQMPK-FQCAT 527
Db 429 FLPHHKNTNSQSYNVVDYV-GTAAPSQKCDLQCGCPAVCINTLLEYRKMDRPPVLSNVK 487
QY 528 KQPIGVCTMNSQVSDCPDIGNVARYLLLRKPDQTEAKAKATMODYRATLEERLEIDLE 587
Db 488 RDPVYITGTAGT-VNDLEILCNFAIFRERREGRGNFVEDAP---KTYWQLOCONI----- 538

```

588 ERLDGRGAPSSGLSSVVDHPTFRLLDPTLAKLEQTTOTPMKVLVETROYKMEGIS 647
 589 EKLAAMGISEGDALEKTLVDPISPKVKRGIDSTVEAELEKFINMK-NNTNPERNIK 597
 648 EATHSMALFDYSGAPPTTINFLVKRTHAAVVDLALSQCCHVYGO----- 695
 598 SVHHLDFACNYYWAPCPVETLKYKSLIVLDICLTS--CMYEDQNPVAVGIVPSHW 655
 696 -----QVEGRNFR-----NQPVLKRRVDFLN-----GGTSTSTIVTISE 734
 656 EKMHQIMNTNKGACFKGATITGGDLKLVHOSMPCDLPDTDAALGCMFAPARMQVR--- 712
 745 GIVSAPNPLDQADAPAGTPEKDLARVSEVVDIHKRRVVSNGCNLSAARLVG 794
 713 -----LAKMLMVKTIKIKRRIIES-NSTG-AESIQAGPMK 747
 745 LAAVAPOROKRVMHJGALPGLKQFHGLLPKCMPPNSKSNPQWFWTL-ORNOP-A 852
 748 PAS QDSY -LVGQYMKFLNALHKLTP-----STKTSALYLMKICQTKNPIL 796
 853 DKLTHEETIA AVKRTVEEVAALINFLPTGIGELAQFYMANILKYCHSOYLINT 911
 747 TVASRPHLEHLYNYKASSQAFREINVDLVDTLSTAKIKLNSILACQDLPYAT 856
 912 L-----TSIIIGARRPPSSVLMHMKDVTSAUDITQAKALLEKTNLPELMTAFT 965
 857 LSTLSVDTOLVABEVP-----HVLGPEVGLSSPD-EYRAK-----VAGR 894
 966 SHLVKAAANO-----KRMVAGISISKYHGAAGNNKRVYQAGN---WSCIINGCKNVC 1014
 895 SATIVOSTIKQAVSINGRKLPIITVPLVVKYTGSGNINVHCAMGYFSGRGVDNRH 954
 1015 P-----LFTFOTRFTIACPGAFICPVTPSSGNRET-TLSDVRCIIVS-GGA 1063
 954 PSVAPKKNVSSMLKRRHVMIPVADRIYKRLVGINSGEPFAAKRSHVQVNLHEDRNP 1014
 1064 MGLAIVAVAVCAKQAHMAFDIMLSITDDE--PLANDLEHLDOLIQLELT----- 1115
 1015 NLPKTVLELVKHISS-----ASTLEDVLYUULPFAVLDEVLSTLSVQAGV 1066
 1116 FWTATLAAVYLLDEKILACIGEPETNLAFNPDSCEPSHDTSNVNLISGNS---ISC 1171
 1067 FWTATLAAVYLLD-----IDDC-----LQVGRPHCLIDG 1099
 1172 STVPLKRPEDDELFLDLSGP 1193
 1100 OSVV -----ELLPSTGVP 1113
 PRELIMINARY: PRT: 1133 AA.
 108904
 108904 PRELIMINARY: PRT: 1133 AA.
 01 MAY 1997 (PREM101: 03, Cited)
 01 MAY 1997 (PREM101: 03, Last sequence update)
 01 MAR 2002 (PREM101: 20, Last annotation update)
 Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 Virus: dsDNA virus; no RNA stage; Herpesviridae;
 Gammatotpressividae; Rhadinovirus.
 NCBI TaxID: 47296;
 11
 SEQUENCE FROM N.A.
 MEDLINE-97094384; PubMed-8949871;
 Moore P.S., Hostolt C., Weiss R.A., Chang Y.,
 "Molecular mimicry of human cytokine and cytokine response pathway
 genes by KSHV."
 Sequence 274.1739.1744(1996).
 12
 SEQUENCE FROM N.A.
 MEDLINE-97121480; PubMed-8962146;
 Russo J.J., Hohenzky R.A., Chien M., Chao J., Yan M., Maddalena D.,

RA Parry J.P., Peruzzi D., Edelman J.S., Chang Y., Moore P.S.;
 RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
 (HHV8)." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Russo J.J., Hohenzky R.A., Chien M., Chao J., Yan M., Maddalena D.,
 Parry J.P., Peruzzi D., Edelman J.S., Chang Y., Moore P.S.;
 Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 075698; AAC57083.1;
 DR InterPro: IPR002114; HPr_Serp_site.
 DR InterPro: IPR001993; Mitoch_carrier.
 DR InterPro: IPR000635; Viral_DNA_bind.
 DR Pfam: PF00747; Viral_DNA_bp_1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 SQ SEQUENCE 1133 AA: 125623 MW: 029ACFAF4C8A5EB7 CRC64:
 Query Match 10.7% Score 675; DH 12; Length 1133;
 Host local similarity 23.6% Pred. No. 56-45;
 Matches 302; Conservative 194; Mismatches 503; Indels 282; Gaps 53;
 11 PRGPGYATCAVEVDLDEETSLAARSTDLALPLKMLVERITTSIAVSGART 70
 18 PGPGCYLAAYLTHNPIDBASLNGYPAKVSILPLHGLTSPSDPLNKAHV--- 73
 71 TGLAGATLKLTTSHFVSFVFGKIVLPSSAPNLTPLNAPPEFGSPVQAPPY 140
 74 KKIDATTAIVKLTSLY--REATVFHNTILFQPLQKGLKLRERELFGSTFVEQOH 141
 131 DCAVETTGAEICTKGLFENILLYLVIALKKAAYKCNVPLHYGDI-----VA 182
 132 KGTL--WSEDAECPQLCAAE-IFMAVITEGEKERL-----YGGKLVPVSQTPVH 180
 183 INHGDVRIELPVOYLFMDVNLVDPDNTIHRSIGELGVPTPRYNGGLCHLHCVI 242
 181 IGBHOAFKILPYDEDLPGSRQEL-----CRYYNDISJYHDSIF 222
 243 APMVAALRVNNTAARGAHLAFDENHETAVLPDITYF-QSSSGTTARFAPRD 301
 223 TGLAQLRKVDVSTVIG-ASERGF-VHDYKIPKLVQKDFQGLASKTDG----- 271
 302 VNSTSKPSPSGCFERRKIASIADDTALAAVLPNGI-YEHTPTDIK-----WMPIG 354
 272 -----STLMVID-SLVAELMSYGLSTFEGPDSCEVLDNTWPFEN 313
 355 MEGTLRLMALASYTARVAGVIGAMFSPNSALYTEDESGTEAKDGAPSPSEKRYO 414
 314 CHTPLAKRLAEVMMHBOALHIGAOVFAANSVYLKLV-AKIPKQNGQDANMTSPYL 371
 415 FASPHILANPQUDRG-----HYLSQSTGSSNTEFSVDYALALIGEGAPLARLEYLE 409
 372 QHGLYLSLEATFKENGASAFKGVPSALDSSS---YTLQHLVASSFPHLARMCYYLO 428
 470 KCDAGAFPGCHG-DALKYVTGTDFDSKIPGSLCEKHTIRPACHTHTYKRLRHMIR-FQUL 527
 429 FLPHKNTSSQSYNVVDY-VTAAPSQKCDLQOGCPRAVINTLFYRMKDRPPVLSNK 487
 528 RQPLGVSTNMSQVSDPDLASNAAYLILKPKDQLEAAKAMQUDYATLELEFLDRO 587
 488 RQPVYITGTAGF-YNDILILGNPALTFRERENGNVEDAP--KYIYWGQONL----- 538
 539 EKLSMGISEGDALEKTLVDPISPKVKRGIDSTVEAELEKFINMK-NNTNPERNIK 597
 588 ERLDGRGAPSSGLSSVVDHPTFRLLDPTLAKLEQTTOTPMKVLVETROYKMEGIS 647
 598 SVHHLDFACNYYWAPCPVETLKYKSLIVLDICLTS--CMYEDQNPVAVGIVPSHW 655
 648 EATHSMALFDYSGAPPTTINFLVKRTHAAVVDLALSQCCHVYGO----- 695
 598 SVHHLDFACNYYWAPCPVETLKYKSLIVLDICLTS--CMYEDQNPVAVGIVPSHW 655
 696 -----QVEGRNFR-----NQPVLKRRVDFLN-----GGTSTSTIVTISE 734
 656 EKMHQIMNTNKGACFKGATITGGDLKLVHOSMPCDLPDTDAALGCMFAPARMQVR--- 712

QY 735 GIVSAFNPITACODAPAGRTPDGLARVSEVLRDLRVKNNRVFSCNCTNLSPAAARALVQ 794
 Db 713 -----IARAMLVPEKTIKRIITFS-NSTG-AESTQGEPMK 747
 QY 795 IASAVQROKRRVDMHCAICLLKQHPGLPPGMPNPSKSNPOFWTLT-OROMP-A 852
 Db 748 IAS-OROMP-IVGCPYMKFNAHKKTLFP-----STKISALYLMKKIQOTTKNPIL 796
 QY 853 DKLTHEEITTA-AVKRTEEYVAALNFILPPTCIGELAQFMANILKCYDSQTLINT 911
 Db 797 PGVSEHILTELCNVKASSQAEEFINVLDLPDTLTSYAKIKINSSILPACQOTQFYATT 856
 QY 912 I-----TSITGARRPPSSVILHMRKDYTSAADETQAKALLEKTENLEMTTAF 965
 Db 857 LSLSPVQLVPAEEF-----HVLGPVGLSSPD-EYRAK-----VAAR 894
 QY 966 SIHLVKAAMNQ-----KPMVVICISISKYHGAAGNNRVQACN--WGLNGCKNVC 1014
 Db 895 SVTIVOSTLKQAVSTNGRLRPILITVPLVNNKYTGSGNGNTNPHCANLGYFSGRGVDRLNR 954
 QY 1015 P-----LFTFKIRPFIHAFKGFICPVTSPSSGNRET-TLSDQVRGIIYSGAM 1064
 Db 955 PESVPPKNNVSSMLRKRVIMFPLVDRIYKRIVGINSGFEFAEAVKRSOVNILEDRO-- 1012
 QY 1065 VOLAIYAIYVKAAGCAQAHMADIMSLITDDE--FLARDIEELHDQIIOTLKT-----P 1116
 Db 1013 -NPNLPKTVLEL---VKFPMWSSCASLSEEDVIYLLGPAVLGDEVLSLLSTVGAGVP 1068
 QY 1117 WIVGCAIHAVKILDEKTTACGDEPTNINLAPNPSRPSHDITSNVINSGSN-----ISGS 1172
 Db 1069 WTARQVAVIOTI-----IDDCB-----LQVGPPEPCLIQGO 1101
 QY 1173 TVPGLKRPEDDELFDLSGIP 1193
 Db 1102 SVV-----EELPSPGVP 1114
 RESULT 15
 Q99226
 ID V99226 PRELIMINARY; FRT; 1103 AA.
 AC Q99226;
 DT 01-JUN-2001 (TREMblrel, 17, Created)
 DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
 DE Single-stranded DNA-binding protein
 GN 6.
 SS murid herpesvirus 4.
 US viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Gammaherpesvirinae.
 NCBI_TaxID=33708;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MHV76;
 RC MEMLINE=21232509; PubMed=11333912;
 RA Metrae A I; Davison A M; Milligan S; Brownstein D G; Allen D J;
 RA Mestrikova J; Davison A J; Nash A A; Stewart J P;
 RT *Analysis of a novel strain of murine gammaherpesvirus reveals a
 RT genomic locus important for acute pathogenesis.*;
 RL J. Virol. 75:5315-5327(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MHV76;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DK EMBL: AF324455; AAK16699.1;
 DK InterPro: IPR002114; HPI_Script;_stc;
 DK InterPro: IPR000531; TonB_boxC;
 DK InterPro: IPR000635; Viral_DNA_bind;
 DK Pfam: PF00747; Viral_DNA_BP_1;
 DR PROSITE: PS00589; PTS_HFR_SER; UNKNOWN_1.
 DR PROSITE: PS00430; TonB_dependent_RFC_J; UNKNOWN_1.
 KW DNA-binding.

FT VARIANT 356 356 S -> G
 SO SEQUENCE 1103 AA: 123272 MW: 056359648627003 CRC64,
 Query Match
 Best local similarity 23.9%; Score 652.5; DB 12; Length 1103,
 Matches 292; Conservative 194; Mismatches 535; Indels 199; Gaps 47;
 QY 1 MENTQKTVIVPPGICGVYACREDLDEETSLAAKSTUSDLALPLMNLIVERTTTS 60
 Db 1 MENTNCSAPLPAGYIYQVLDDEPLEASLSTNFSAKACLPVLTGLVEGFNF 60
 QY 61 SLAAVSGARTTGLAGAGITLKTSTSHFPPSVFVPHGCKHVPSSAAPHNIPRACNAAREP 120
 Db 61 NYK-----APCQKIMESLTVKPT--KYPNHAIVFHSNKRIPPIFHGRGLFQICEAIRLMP 114
 QY 121 GTSRCGGPPYDCAVETTGAEICTRLGLEPBNITLYLVYALFEAVFMCNVFLHYGLDI 180
 Db 115 GFSSFKPDASKGQMKSS-----THLPQNEPKYIGVVVAHSEFKRLICGLIAMSALPMKT 169
 QY 181 VHNHGDVIRIPLPVQLMPDVNRLVPPDFNTHRSIGCGFYPTPPYNTGICHLIHOC 240
 Db 170 VIINMLCKKVPPLHDIRLSSD--QLL-----PEYDENYSKIMYES 208
 QY 241 VIAPMAVALRVNVTAVAGCAHIAFDENHEGAVPPDITTYTFQSSSGCTTARGARRN 300
 Db 209 YVTNVAQAVRIHDVGLIALDRAVTHQYKAKLASIKTY----- 249
 QY 301 DNSTSKPSPSGFEERLASIMAAQTALIAEYFNNGIYEETP-TDICKPMFTMEGTL 359
 Db 250 ----PAENCGAGNILEYAVIDAAVSELAWSHGLAFLEAPQEESSILQNVYQVPMF--SHDAT 403
 QY 360 P--RLNALGSIYARAVAGVIGAMVFSPNSALYLTVEEDSCMTEAKDGGPSF-NRERYORA 416
 Db 304 PGRIRIKALDEMAKMAIHHAQVLSANSVLYTVKVGQSA-GRNVGKQKDDMELSKRYMOH 462
 QY 417 G-PHIAANPQTBDRGHVLSSTGSS--NTERSVYDALICGFGAPILAKLLFYLEKCD 472
 Db 363 GLGHL--NVQTDENNTVFPVPGVPCMGNSQYTVHNLAAASMSPHILAKLCYVLYOMCQ 420
 QY 473 AG--AFTGGHGDALKVYVGTGTPDSEIPECSLEKHTPPVVAHTVHPLKQMPF-PVQATP 529
 Db 421 NGRLASTPGCTTVPATVTTANTSM-CDLCCGDLPAVCLHTLFFRLRDRKFOVLATQKKD 479
 QY 530 PIGVGTNMSQYSDCDPLGNAPAYLLKPKGQUTAAKAIMQUTYKAILKLPIDLEOKR 589
 Db 480 PYVVTG-IGAPHNDMELIDNGFSF--KERDDQ-----ENGQRYSVQUNANL-TER 527
 QY 590 LIDRGAPCSSEGLSSVYDHPTEFRRLDPLRLRIQTTQPKVIVYTR-DYKIRGLISF 648
 Db 528 LDEIGVTASTPA-CQLITDVDSFLTVFKKIDAVDEETIRFNDGLVRNNINRK--ESVKS 584
 QY 649 ATHSMALTFDPYSGAFCPITNPLVKTTHLAVVODIALSQC-----HCVFYQVVGKRN 702
 Db 585 IAHVILQDQNTWMLPPCPRIHNIIFTRSPFTIIDMSFPLCVIHHTENPLNYG--TVSTWM 642
 QY 703 RNQFQVPLRRRVLDLNGFGFTSKSLIYVLSKSPSAPNPIIGQAPAGRPVGGIARVS 762
 Db 643 SNHFQIMNSKFSIMFDKLLTCSDMRVVHTE--TSSDMENIQAADLPKQPCAI-PLSPAQ 699
 QY 763 VEVTRIDIRKNNRVFSCNCTNLSEFAARALVGLASYYQGFQKRVDMHCAICLLKQHPG 822
 Db 700 IFAPKALKIKNNILFSSNSS--TESLSASFVAGCAEN-----PTACPYMHFPILOLHK 751
 QY 823 ILFPKMPNPSKSPNUPFW-TLQORNOP-----ADKLTHEEITTAIVKRFTEEYVA 875
 Db 752 QLPF-----YTKGAYYIMQTFMSSKLIPOAGSTNAKLKS-EFISYIMTSSIAHDE-- 801
 QY 876 INFIMLPPTCIGELAQFMANILKCYDSQTLINTISI-----ITGAP 921
 Db 802 VAVLDVYPTTLMAKAKQRLNAILKLCGTVQYATTINFLQETLSMPSLDYPIHVG-- 858
 QY 922 PRDSSVILHMRKDYTSAADETQAKALLEKTENILEMTTAFSTHILVKAAMNORPMV 981

```

00 859 PIVV- Q011YMDMTHDOSAIVLOSCHRD-----PAVCKMRPIVT 899
0Y 982 IATISISKYHGAUNNREVEAGNNSGUNGKNGCPLFTDR-----TRR 1024
00 900 IPMANVKYSNGNNSIFUSGN MGFMGKV-----DRKLLDIPRMKHVNTSMRRR 952
0Y 1025 F11ACPRGCEIYVIGSSGNRETTISDOVR-----GIVSGAMVOIAIYAVVRAVGAR 1080
00 954 YAFASHPTEST- LTQPGKIVRTYQLEKTRKI IHDILASGNAPE- NLVLELVKCLGHE 1009
0Y 1081 AQHMAFDOMLSTLDGFFLAKGLFELHDJITOTETPTVEGAL- EAVKILDEKTAGDG 1138
00 1010 CATLISD- DVEFYAGFSPHVNVDVMEVLETLKASGCAFSESSVOALAPTSBOTS 1063
0Y 1159 ETPTNLAENFDSPEPSHOTT 1158
00 1064 ESPVEVTDNITSLPEISVT 1083

```

Search completed: March 28, 2003, 14:48:02
 Elapsed time: 118 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 13:31:04 ; Search time 30 seconds

(without alignments)
1663.201 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTOKTVTPGTPLGYVA

Scoring table: BLASTUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: SwissProt_40*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1204	1 DNB1_VZVD	P09246 varicella-z
2	3620.5	57.5	1209	1 DNB1_HSVB	P28932 equine hefp
3	3054	48.5	1196	1 DNB1_HSV1K	P17470 herpes simp
4	3049	48.4	1196	1 DNB1_HSV1L	P04296 herpes simp
5	3047	48.4	1196	1 DNB1_HSV1L	P17469 herpes simp
6	3015.5	47.9	1196	1 DNB1_HSV2H	P89452 herpes simp
7	3013	47.9	1197	1 DNB1_HSV2	P36384 herpes simp
8	2933	46.6	1186	1 DNB1_HSVB2	P12637 bovine hefp
9	1094.5	17.4	375	1 DNB1_HSVB1	P03444 equine hefp
10	669	10.6	1128	1 DNB1_EBV	P03227 Epstein-Bar
11	604	9.6	1128	1 DNB1_HSVSA	P24910 herpesvirus
12	492.5	7.8	1132	1 DNB1_HSV6U	P52338 human herpe
13	490.5	7.8	1132	1 DNB1_HSV6Z	P52338 human herpe
14	483	7.7	1131	1 DNB1_HSV7J	P23239 human herpe
15	479.5	7.6	1160	1 DNB1_HSV7C	P13215 simian cyto
16	477	7.6	1135	1 DNB1_HSV7A	P17147 human cytom
17	473	7.5	1193	1 DNB1_HSV7A	P09117 herpesvirus
18	412	6.5	1191	1 DNB1_HSV7A	P30672 murine cyto
19	421	1.9	1066	1 DNB1_HSV7A	P09113 homo sapien
20	415	1.8	833	1 DNB1_HSV7A	P18401 gallus gall
21	414	1.8	833	1 DNB1_HSV7A	P18401 gallus gall
22	414	1.8	833	1 DNB1_HSV7A	P18401 gallus gall
23	413.5	1.8	833	1 DNB1_HSV7A	P18401 gallus gall
24	413	1.8	833	1 DNB1_HSV7A	P18401 gallus gall
25	410	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
26	409	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
27	408	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
28	408	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
29	407	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
30	406.5	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
31	406	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
32	406	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall
33	406	1.7	2569	1 DNB1_HSV7A	P18401 gallus gall

ALIGNMENTS

34	105.5	1.7	1016	1	UVRA_DEIRA	046577 deionococcus
35	105.5	1.7	2114	1	MY9B_MOUSE	094906 mus musculu
36	105.5	1.7	4540	1	DYHC_PARTE	027171 parametium
37	102.5	1.6	1902	1	P2P_LACTIC	P15293 lactococcus
38	102	1.6	1038	1	V120_VZVD	P09277 varicella-z
39	102	1.6	1275	1	RFBG_MYXA	050864 myxococcus
40	102	1.6	4590	1	FATH_HUMAN	014317 homo sapien
41	101.5	1.6	1478	1	BUD3_ASHGO	094161 ashbya goss
42	101.5	1.6	1502	1	PTP_LACTIC	P16271 lactococcus
43	101	1.6	1348	1	CV43_TRYBH	099280 trypanosoma
44	101	1.6	1348	1	VCR2_COTUA	P52583 coturnix co
45	101	1.6	3649	1	ACVS_MOCIA	P27743 nocardia la

RESULT 1
ID DNB1_VZVD STANDARD: PRT: 1204 AA
AC P09246:

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Major DNA-binding protein.

GN 29.

OS Varicella-zoster virus (strain Dumas) (VZV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirus.

OX NCBI_TaxID=10338;

RP SEQUENCE FROM N.A.

RX MEDLINE=86306657; PubMed=3018124;

RA Davidson A.J., Scott J.E.;

RT "The complete DNA sequence of varicella-zoster virus.";

RL J. Gen. Virol. 67:1759-1816 (1986).

CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA

CC REPLICATION.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: X04370; CAA27912.1; -

DR PIR: C27214; DMB829.

DR InterPro: IPR00635; Viral DNA_bind.

DR Pfam: PF00747; Viral DNA_bp: 1.

KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

FT ZN-FING 497 510

SO SEQUENCE 1204 AA, 132139 MW, D2F6F65DC0C674 CRC64;

Query Match 100.0%; Score 6294; PA 1; Length 1204.

Best Local Similarity 100.0%; Pred. No 0;

Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MENTOKTVTPGTPLGYVAACRVERNDLFPISFLAARSTDSNALLPLMRNITVVKPTFS	60
DB	1	MENTOKTVTPGTPLGYVAACRVERNDLFPISFLAARSTDSNALLPLMRNITVVKPTFS	60
QY	61	SLAVVSGARTTGLAGAGITLKTTSHPFSVFHFGCKHVLPSNAAPLRTAGNAAAREP	120
DB	61	SLAVVSGARTTGLAGAGITLKTTSHPFSVFHFGCKHVLPSNAAPLRTAGNAAAREP	120
QY	121	GFSCGCPVVDCAVEETGAEICTRLGLEPNTILYLVVLTALFKAVFNCNPLHYGGLDI	180
DB	121	GFSCGCPVVDCAVEETGAEICTRLGLEPNTILYLVVLTALFKAVFNCNPLHYGGLDI	180


```

QY 419 HLAAPQTDGRGHV-----SSOSTSSNTEFSVDYALALICGAPILARLLFLERCD 472
DB 418 YVACQPDQDKGRVLPHPADQOAAPIANSNOEFLSDYIALACGFCQIILARLLFLERCD 477
QY 473 AGATGT-GHGALAKVSTFSELPSCLEKHTPRVCANTHYHRLQRMPPFGQATROP1 531
DB 478 AGTGGGRNETALRYLANTLESDDVPCGCLNOATBPACAHLLHRLRORLPFGGAVPRAPI 537
QY 532 GVFQGMNSQYSCDPIGWAPYPLLRKPGDQTEAKATMODYRATLRPLFIDLRORHL 591
DB 538 GIFTMNSAYSDCDYLGWYASYGALKRPND-NEAKSIMODYRATMELELLEQAKLI 596
QY 592 DR-----GAPCSSEGLSSVVDHPFRRLDILRARIEDOTTQFKVLYETRDYKIREGL 646
DB 597 DKETLAQSPSCAP--TSVHDAQSFICGLSNIKDTICGAAGQFRLTLEADPFIRESL 654
QY 647 SEATHSMALTDPRYSGACPTINFLVKRTHLAVODLALSGCHCYFGQGVGGRNFRNOF 706
DB 655 ADANHTMSISLDPYSSSFCPTSTFLARTVFAVLDDLVLSCHCLFYGQSVGGRNFRNOF 714
QY 707 QPVLRRRFRVDLPNGGFISTRSTVTTLSECPVSAFPTLGDAPAGRTFGDLARVSVEYI 766
DB 715 QPVLRRRRLDMLNGGFIITAKTIVTVVSDSGVLAAPDLTPRASEPTPKDYDGMARSMELV 774
QY 767 RDIYKKNVFSNGCTNI-SEARARLVGLASAYQROEKRVDMELHGALELLKQFGLLEP 826
DB 775 RDIYKKNRVLFSNGCANNSFAAPAVACMASAYBRDKGNILNGAVFLVQYQGVLEP 834
QY 827 KGMFENKSPFQUMFTLQKUMPAOKLTHEITTAIVKRFTEEVAAININLPICI 886
DB 835 KGPFGIDTPPQWFTLQKUMPAOKLTHEITTAIVKRFTEEVAAININLPICI 894
QY 887 GELAQFYMANLLKYCDHSDVYINNTSTITGARRRDPSSVLMIRKDVTSAADETVQ 946
DB 895 GELAQFYMANLLKYCDHSDVYINNTSTITGARRRDPSSVLMIRKDVTSAADETVQ 954
QY 947 KALLEKTENPELTTATSTHLYRAMNORPAAVYLGISIKYHGAAGNNRVFOAGNNSS 1006
DB 955 QEVLRQSLNSNAAMTGTSTINMAYVMDORPVVYLGISIKYHGAAGNNRVFOAGNNSS 1014
QY 1007 LGGKKNVCLPTEFTRTRFIACPRGFGICPVTGSSNRETTLSDOYKGIIVSGAMV 1066
DB 1015 LGGKKNVCLPTEFTRTRFIACPRGFGICPVTGSSNRETTLSDOYKGIIVSGAMV 1074
QY 1067 LAIYAVYRAVCAQAQMAFDWLSLTDDEPLARDLELHQIITYLETPWTVEGALP 1126
DB 1075 TAVFAAVYLAHAGARTQHLAVLWLTGLVDEFLAASLALNATVYDQF-GEMSVFAAUELY 1133
QY 1127 KILDEKTAG--DGETPTNLAFNFDSC---EPSHDTTSNVLINISGNSISGTVGCLRP 1180
DB 1134 KMFRAQTTAGVAAQEG---AFDQACVCDTTOOSTSA---FNGCLAMAAPAGOKRS 1185
QY 1181 PEDELEPLSGIPIKHGNITMEM 1203
DB 1186 LPDILTFQW:APPEKSGILTFDM 1208

```

```

RX MEDLINE-88179536; PubMed=2833010;
RA Gao M., Bouchev J., Curtin K., Knipe D.M.;
RT "Genetic identification of a portion of the herpes simplex virus ICP8
protein required for DNA-binding."
RL Virology 163:319-329(1988).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
REPLICATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
(PREREPLICATIVE SITES). AS VIRAL DNA REPLICATION PROCEEDS, IT
MIGRATES TO GLOBLAR INTRANUCLEAR STRUCTURES (REPLICATION
COMPARTMENTS).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M20165; AAA5793.1; -.
DR PIR, A28601; DNEBS.
DR InterPro, IPR000635; Viral_DNA_bind.
DR Pfam, PF00747; Viral_DNA_bp.1.
FT DNA-binding; DNA replication, zinc finger, Nuclear protein.
FT ZN-FING 499 512
SO SEQUENCE 1196 AA; 128314 MW; 0D0010A5FDPA464 Cpw64;

```

Query Match 48.5%; Score 3054; DB 1; Length 1196;
Best Local Similarity 49.8%; Pred. No. 5; 3e-228;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```

QY 1 MENTOKTYV---VTPGRLGYV--ACRVEDLDELISFLAASSTSDLLPLMRUTVE 55
DB 1 MENTOKTYV---VTPGRLGYV--ACRVEDLDELISFLAASSTSDLLPLMRUTVE 58
QY 56 KFTSSLAIVSGARTTGLAGAGITLKLITSHFVSFVFRGCKHYVLPSSAANLPTRACNA 115
DB 59 SGEFANVANVVSRTTGLGTAIVSLKLTSHSSSVYVHGGRHIDPSTQALNTRLCER 118
QY 116 ABERGFSSCGRPVDGAVETTGAEICTRLGLEPENTILYLVVTLAKVEMCVNELHY 175
DB 119 ARRHGFSYTPRPDGLKHEITGEALCERLGLDPBALILYLVTEBQKVAVCINNTPLH 178
QY 176 GGLDIYHINHSVITRLPEFYQLEMPDVNRKLVPRDPFNTHHKSIGSEYVPTFEYNTGLCH 235
DB 179 GGSDDVTTIGAEVNHIPVPLDQLEMPDVSRYLAEPFNHKSIGSEYVPTFEYNTGLCH 238
QY 236 LTHDCVIAAPMAVALRVKNTAVAKGAHLAFDHNHCYAVLPPIYIYFQSSSSGTTIAR 295
DB 239 LTFEAVVGPAAVALRCRNVDAVAAAHILAPDENEGALLPADITTAEEAASGG-KTPR 296
QY 296 GARRDNVNSTKPSFGSEERKLASIMADTLALHEVINTGITYETPTDLEMPEFTGM 355
DB 297 GGR-----DGGKGRPAGGEGEORLASVMAQDAALALESIYSMAVFPDEPPTDLSAMPICBO 351
QY 356 EETLRLNALSGYTA RVAGVIGAMVFPNSALYLTVEEDSGMTEKDGPGSSENFYGF 415
DB 352 DTAARANNAVGVILARRAGLVGAMVFTNSALHILEVDAGADPRDHSK-PSFYREFLY 410
QY 416 AGPHLAANDPDRDGHV-----SSOSTSSNTEFSVDYALALICGAPILARLLFLY 468
DB 411 PCTHVAANQYDVRHGVVGFEGRPAPLVGGTO-EFAGEHIALMLCGFSFALLAKKFLYL 469
QY 469 EECDDGALFGGNG-DALKVVTGTFSSEITPCSLCEKHTPRVCANTHYHRLQRMPPFGQAT 527
DB 470 EECDDGALFGGNG-FRVLVADSNOTVPCNLTCTDTRACVHTLMRLKRNHPRFASDA 529
QY 528 RQPIGVFGTMSQYSDCDPLGNYAPYPLLRKPGDQTEAKATMODYRATLRPLFIDLEU 587

```

```

10b 540 KCAICVPCSTMSVSNVTVICNVAAPSAIKR ADGSEIARTIMOTYPAATFRMALEET 588
11b 588 ERLIDRAVMSSELSIVVHPFRRLIDTIRATIEDTTPMVLVETPRYKIPRELS 647
12b 589 LQYVQAVPTAMKLETTITNREALITYVNNRVQVDEVEFLMNLVGPFRKFRDIA 648
13b 648 PAFHSMALTEPVYSCAPCTINFTKRTTHAVVQIALISQCHVYVGOVGRNPNRPO 707
14b 649 EANIMASLTIDPYACGPELLOLTKRSNLAVYODALSQCHVYVGOVGRNPNRPO 708
15b 708 PVLARRVVDLNGGFTSRSTITVTLSEG-PVAPNPPTLGQAPARTEDGLARVSEVI 766
16b 709 PVLARRVVDLNGGFTSRSTITVTLSEG-PVAPNPPTLGQAPARTEDGLARVSEVI 768
17b 767 RDIYKKNVSVSNCTNLSFAAFATVILASVYQCKRKMVLHSLFELKQFRLTFP 826
18b 769 KRLVYKSVLTLGASANSSEAKAVASVLOQYQKPKRVDLLGPTFLKQFRLTFP 828
19b 827 KQMPVNSKSPNQQWVLLQHNQMAIKIETHETTTAAVKKFTFEVAAINFLPTTC 886
20b 829 NKRPVNSQPNQWVLLQHNQMAIKIETHETTTAAVKKFTFEVAAINFLPTTC 888
21b 887 GELAGVYVANIILKYGCHSVYLITLSTTSAPRPSPSVLHMPKRVTSAAITETQA 946
22b 889 SELAMVYVANIILKYGCHSVYLITLSTTSAPRPSPSVLHMPKRVTSAAITETQA 945
23b 947 KALLEKTENELMTATSTHLYVAAANNQRPVVLGISTSKYHGAANNVYVGAANNSG 1006
24b 946 KALMAYVAVHIEVAMTSMFASNLPLPVMAAPRVVLGLSTSKYHGAANNVYVGAANNSG 1005
25b 1007 LMGKNGVCLTTPKPKFPIACHPGCTGVYIGSSGNKRETTISDOVKGIIVSCGAWQ 1066
26b 1006 LMGKNGVCLTTPKPKFPIACHPGCTGVYIGSSGNKRETTISDOVKGIIVSCGAWQ 1065
27b 1067 LAIYATVAVAAVAAQVIAAFDMLSLTDEFLARDEELHQQIOTLEPTVEGAL 1123
28b 1066 SSVAVAIYKSLAPRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1125
29b 1124 EAVKILDEKTTAGGCTPTNLANFDD--SCPSHDTTSNVLNIGSNGSTVGLKRP 1180
30b 1126 HEAVAVSGLANAP--VNFVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1173
31b 1181 PEDEDELPSLPSLPSLPSLPSLPSLPSLPSLPSLPSLPSLPSLPSLPSLPSL 1204
32b 1174 FPGDHPRG EGPVPRKKKGLTLLM 1195

```

```

RT type 1 containing the genes for DNA polymerase and the major DNA
RT binding protein.
RL Nucleic Acids Res. 13:8143-8163(1985).
RN [3]
RP SEQUENCE OF 1062-1196 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hamersschmidt W., Contrals F., Mankertz J., Huhk H.-J., Paul G.,
RA Ludwig H.
RT *Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).
RL Virology 165:406-418(1988).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -2- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC (PREREPLICATIVE SITES) AS VIRAL DNA REPLICATION PROCEEDS, IT
CC MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC COMPARTMENTS).
CC -3- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC This SwissProt entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Message by and for commercial
CC entities requires a license agreement (see http://www.isb.slb.ch/announcement/
CC or send an email to license@isb.slb.ch)
CC
CC EMBL: D10879; BA01675.1;
CC EMBL: X14112; CA3232.1;
CC EMBL: X03181; CA26940.1;
CC EMBL: M21631; AAA5787.1;
CC PIR: A03790; DNEB1.
CC PIR: B30085; B30085.
CC InterPro: IPR00635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_dp_1.
DR KJ: DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN-FING 499 512
FT SEQUENCE 1196 AA; 128449 MW; 453794b2E5B90E9 CRC64;

```

```

Query Match 48.4% Score 1049, DM 1, length 1196,
Host local Similarity 49.8%; Pred. No. 1,362,227;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

```



```

DB 952 LDAHFGAMTSMFASCNLKPVAARPMVVLGLSTISKYMGAGDRPVQACNNASLLGCKN 1011
QY 1013 VCPLETFDTRRFTIACPGRGFTCPVTGSSGNRETTISDQVKGITVSGAMVOLAIVAT 1072
DB 1012 ACPLILFDTRKRFVLACPRAGFVCAASSLGGGAHSHSLCEFOJLKGIIAAGCAVAASSVFA 1071
QY 1073 VRAVGAQAQMAFPDWMISLTDDEFLARDLELDQIOTLETPMTVEGAL-----EAVKI 1128
DB 1072 TKSLSGPRGQOQIEDWMLALLEDEYLSEEMMETTRALERGHEGEMSTDALAEVAHEAEAL 1131
QY 1129 LDEKTTAGDGETPTNLANFNPDSCEPSHDTTSNVLNIGSNIQSTVPGLKRPEDDELDP 1188
DB 1132 VSQIACAGH-----VNFNGDFGDHDDHAASFGGLAA--AGAAVAKKRAAFHGDPPG 1182
QY 1189 LSGPIKIGNITMEM 1203
DB 1183 -EGPEKK-DITLDM 1195

```

RESULT 7

```

DDBI_HSV2 STANDARD: PRT: 1197 AA.
ID DDBI_HSV2
AC P36384
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).
GN DBP OR 0129 OR ICP8.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310.
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=93228441; PubMed=8385914;
RA Toh Y., Liu Y., Tanaka S., Mori R.;
RT Nucleotide sequence of the major DNA-binding protein gene of herpes
RT simplex virus type 2 and a comparison with the type 1.
RL Arch. Virol. 129:183-196(1993).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_bp.
DR DNA-binding: DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 499 512 C2HC-TYPE.
FT DOMAIN 1189 1197 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT SEQUENCE 1197 AA: 128412 MW: 1576858865DBF CRC64:
SQ

```

```

Query Match 47.98, Score 3013, DB 1; Length 1197;
Host Local Similarity 49.08; Prev. No. 7.9e 2.5,
Matches 594; Conservative 213; Mismatches 367, Indels 40; Gaps 16;

```

```

QY 4 TQKTVPTGTGLGVY--ACREDEDELEISFLAARSTDSLALPLMRNLTVEFTSS 61
DB 7 TTTIVKVPQMGVYVYACAPARGLLE--LSLISARSDAIVAAAPLIVGLTSSGFAN 64
QY 62 LAVVSGARTTGLACAGITLKITSHYPSVVFHCGKHLVPPSSAAPNLTRACNAARFPG 121
DB 65 VAAVAVGGITGLGTAVSLKMPSHYSPSYVVFHGRILAVSTQAPNLTRICERARRHF 124
QY 122 FSRQCGPPVAVAFETGACITRIGLEPENTILVLTALPKFAVFNQNVLAHGGIDIV 181
DB 125 FSOYARPCDLKHETTDALCHRLGADPDRAILVLTITGCRFAVQISNFTLHCGMKV 184
QY 182 HINHGVIIRLPFVQLMPDVNRLVPDPENTIHRSIGEGVFYPTPYNTGLCHLIDCV 241
DB 185 TIGDAEVHRIIPVYPLDMHMDPSHVIADPFNCNHRISIGENFNYPDLPFNRLARLFEAV 244

```

```

QY 242 LADMAVALRYNVTAAVARGAHLAFDENHCAVLPDITVTFQSSSSGTTAKGARND 301
DB 245 VGPAAVALRARNDAVAAAHLAFDENHGAALPDITITAEASG--KPRGAR--- 299
QY 302 VNSTSKPSPGGERLASSIMADTALHAIVTNGTGYEETPPDIKEMPEICMEGTLPR 361
DB 300 --DAGKKGPRGGEORLASMAQDAALESIVSMVFEDEPPDITWPLLEQGFYPAAR 357
QY 362 LNALGSYARVAVIGAMVFPNSALYTEVEDSGMTEAKDGPSPFNFFYOFAPHILA 421
DB 358 AGAVGAYLARAAAGLVCAWFSTNSALHLEVEDAGPADPDHKK--FSYFTFLVPGTHVA 416
QY 422 ANQOTDRDGHV-----SSQSTGSSNTEFSDVYALICGFCAPILARLFLYERCDAG 474
DB 417 ANQDRBHHVYVGYEGRPTAPLVGTO--EFAGEHLAMCGFSPALLAKMLFYLERCDG 475
QY 475 AFTGHHG-DALKYVTGTFPSEIQCCEKITRVCANTHTVHRLROMPRGATKQPIV 533
DB 476 VIVGQEMDVFRYVADSGQDVPCNCTPETRACAHITLMRLRAHPKFAASARAGIV 535
QY 534 FGTMSQYSDCDEPLGAYAYLILRKPDOTEAKATMDOTYPATLEFLDEQERILDR 593
DB 536 FGTMSAYSDCQVLGNVAAFSALKR ADGSENTRTIOMETYRATERVAALEALQYVD 594
QY 594 GAPCSSEGLSVIVDPTRRLIDTLARIFQTTQFMKVIYVTRDYKI RRGISEATISM 653
DB 595 AVPTALGRLTETIIGNREALHTVNNIKOLVREDEVDLMNLEGRNFKFRDGLAEINIM 654
QY 654 ALTFDPSGAFCDPTNLELKRTHLAVODLALSQCVEYGVUVEGRNFRNFOPLARR 713
DB 655 SLIDPYTCGPCLLOLLARRSNLAVYODLALSQCAGVACGVBGRNFRNFOPLARR 714
QY 714 FVDLEFNGFISRTSITVTLSEG--PVSAHPITIGQAPARTDGDGLARSEVETIRYK 772
DB 715 VMDLEFNGLSAKTITVALSEGAICAPSLTAGQTPASSSEGDVARITLFPKELRX 774
QY 773 NRVEFSGNCTNLSEARARIVGLASAYOROKRVIMAGALFILLKOPHLLFPPGMPN 832
DB 775 SHVLFAASANASEAKAKAVASLOSAYQKPKRVDDILGFLPLKQFNAVIFPNKPRG 834
QY 833 SKSPDQWMTLORONMPADKTHEITITAAVRFTEYAAINFINLPPTCIGLAOF 892
DB 835 SNQPDQWMTALQNNOLPARLSREDIETAFIKFSLDYDAINFINAPNVSLAMV 894
QY 893 YMANLILYCDHSDYLINTLTSIITGARRPDPSSVLMIRKDVTSADIETQAKALLEX 952
DB 895 YMANOILRYCDHSTYFINTLTAVIGSRPSPVQAAMAPQ---GAGLEAGARLMD 951
QY 953 TENLPELMTATSTHILVRAAMNORPVYVLGISISKYHAGACNNRVPOAGNNSLGCCKN 1012
DB 952 LDAHFGAMTSMFASCNLKPVAARPMVVLGLSTISKYMGAGDRPVQACNNASLLGCKN 1011
QY 1013 VCPLETFDTRRFTIACPGRGFTCPVTGSSGNRETTISDQVKGITVSGAMVOLAIVAT 1072
DB 1012 ACPLILFDTRKRFVLACPRAGFVCAASSLGGGAHSHSLCEFOJLKGIIAAGCAVAASSVFA 1071
QY 1073 VRAVGAQAQMAFPDWMISLTDDEFLARDLELDQIOTLETPMTVEGAL-----EAVKI 1128
DB 1072 TKSLSGPRGQOQIEDWMLALLEDEYLSEEMMETTRALERGHEGEMSTDALAEVAHEAEAL 1131
QY 1129 LDEKTTAGDGETPTNLANFNPDSCEPSHDTTSNVLNIGSNIQSTVPGLKRPEDDELDP 1188
DB 1132 VSQIACAGH-----VNFNGDFGDHDDHAASFGGLAA--AGAAVAKKRAAFHGDPPG 1182
QY 1189 LSGPIKIGNITMEM 1203
DB 1184 -EGPEKK-DITLDM 1196

```

RESULT 8

```

DDBI_HSV2 STANDARD: PRT: 1186 AA.
ID DDBI_HSV2
AC P12639;

```



```

0S Equine herpesvirus type 1 (isolate HVS25A) (EHV-1).
0R Viruses: dsDNA viruses, no RNA stage, Herpesviridae;
0C Alphaherpesvirinae, Varicellovirus.
0X NCBI_taxid=10327;
0K 111
0R SEQUENCE FROM N.A.
0R MEDLINE=94106109; PubMed=8279122;
0R Bell C W, Whalley J M.;
0R "Herpesvirus icp18.5 and DNA-binding protein genes are conserved in
0R equine herpesvirus-1."
0R Virus Genes 7:219-228(1991).
0C -1 FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
0C REPLICATION.
0C -1 SUBCELLULAR LOCATION: Nuclear (Probable).
0C -1 SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
0C FAMILY.
0C -----
0C This SWISS-PROT entry is copyright. It is produced through a collaboration
0C between the Swiss Institute of Bioinformatics and the EMBL outstation -
0C the European Bioinformatics Institute. There are no restrictions on ways
0C use by non-profit institutions as long as its content is in no way
0C modified and this statement is not removed. Usage by and for commercial
0C entities requires a license agreement (see http://www.isb-sib.ch/announce/
0C or send an email to license@isb-sib.ch).
0C -----
0R EMBL: D13930; BA003033.1; -.
0R PIR: J00846; J00846.
0R InterPro: IPR000635; Viral_DNA_bind.
0R Pfam: PF00747; viral_DNA_bp.1
0R DNA-binding; DNA replication; Nuclear protein.
0R NON_TER 1
0S SEQUENCE 375 AA: 40309 MW: ECP327925EBF999B CRC64.

Query Match 17.4%; Score 1094.5; DB 1; Length 375;
Best Local Similarity 55.5%; Pred. No. 4.9e-77;
Matches 212; Conservative 57; Mismatches 98; Indels 15; Gaps 5;

QY 828 GMFEKSKSFNPQWMTLLQGRNMPADKLTHEITITIAAVKRFTEYAAINFINDEPCTIG 887
DB 2 GHPPGIDTPNPQWFWPTLLQGRNMPARKLSKEDIEITITIAKRPDSYSAINFINLPNNIG 61
QY 888 ELAOFYMANLILKYODHSQYLINTTISITGARRRDRSSVYHWIRKQVTSAADETQAK 947
DB 62 ELAOFYFANLVLYKODHSQYFINGELTAIVGSRRRDPAAYLAIVLMDRIINGAAVDEPAAQ 121
QY 948 ALLEKTEMLPELWTTATFSTHLYRAAMNORPVVLGISISKYGAAGNNRRVQAQNMWGL 1007
DB 122 EYLQRLGSPNPAAMTSTETSTNNRYVMQDQRPVYVIGLSISKYSGAGNNRRVQAQNMWGL 181
QY 1008 NQGRKVVCLFTFPDPFRPFPIANPFGGCTCPVTPSPSSQRETTLSQVQVGIIVSGAMQL 1067
DB 182 NQGRKVVCLPMATDKRRKRYVLACPKVGTCEAGQFGCTGRENLTLSQVQVGIIVSGGPMQOT 241
QY 1068 AIYATVAVAVGAPAHMFMETDMLSTDEEFLLARDEELHDQIIVLEFPVTEAGLEAVK 1127
DB 242 AVEFAVLVHLNGARCTQHLVAVDDWIGVDEDFLAASIDALNATVDDGF.GEMSVTEAAGDELVK 300
QY 1128 IIDEKTAC---DGETPPNLAAPNPSC---DPSHTTNSVNLISGSNISGSIVPOLKRP 1181
DB 301 NNEAQTITAGAAVAGEG---AFVQACVGVDPVQVSTSA---FNGILMAAAPAGQKRSL 352
QY 1182 EDELEFELSGIPKIGNITMEM 1203
DB 353 PDDILIFQMGAPPEKKSGILIPDM 374

RESULT 10
DB1_EBV STANDARD: PRT: 1128 AA.
DB1_P33227:
DB1_21-JUL-1986 (Rel. 01, Created)
DB1_21-JUL-1986 (Rel. 01, Last sequence update)
DB1_16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Major DNA-binding protein.
UN BALE2.
UN Epstein-Barr virus (strain B95-8) (Human herpesvirus 4) .
CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC Gammaherpesvirinae; Lymphocryptovirus.
CX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RA MDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Watfoll B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984) .
CC -I- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable) .
CC -I- SIMILARITY: BELONGS TO THE HERPESVIRUS DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)."
CC -----
CC EMBL; V01555; CAA24808.1; -.
CC PIR; A03791; Q0BEA7.
CC PIR; S33057; S33057.
CC InterPro; IPR000635; Viral_DNA_bind.
CC Pfam; PF00747; Viral_DNA_bp; 1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein;
CC Early protein.
CC KW SEQUENCE 1128 AA; 123121 MW; 283D5258031A16D6 CRC64;
CC -----
QY Query Match 10.6%; Score 669; DB 1; Length 1128;
QY Best local similarity 23.2%; Pred. No. 2.5e-43;
QY Matches 292; Conservative 186; Mismatches 537; Indels 246; Gaps 45;
QY
Db 13 GRLGVYACRVEDDLEISFLAARSTDSLDALFLMNLIVKFTISLAVSGARTTG 72
Db 18 GPGGYIYPLATYLRVATLGSTGAGHRCGLTVLLGIIYEPFSIN--VAMLIHRP 74
QY 73 LAGAGITLKLTTSHRYSPVYFPHGKHNLPSSAAPNLTACNAAREHRRGFS-----RCQ 126
Db 75 DPNCGL--LVAISYHRIYVFNHNAWPRPIFEGGELALGCTREVEGVDAYSALPRES 131
QY 127 GPVYGAVEITGALCTLGLEPENTLIYLVATLAFKAVMCNVEFLHVGGLDYHINMG 186
Db 132 SKPGGFPE-----GLDPSAYLDAVAITFAFKRLISGNLVAIPSLKQYAVGOS 181
QY 187 DVIRILFEVOLFMEDVNRILVPRPENTHRSIGEFYVPIFYNTGLCHLHDCVIAEMA 246
Db 182 ASVRVPLDKVEFEGVPUL-----PQFYNSDLRCNHEALYTGLA 222
QY 247 VALRRVNTAVARGAHLAFDENHEGAVLPDITTYTQSSSGCTTTAKCARNDVNST 306
Db 223 QALRRRRGKL-----VELLEKQSLQDQAKVAKVLPKEFPAST 261
QY 307 KPSPGGGERRLASIMADT-ALHAIEVFNTGIYE--ETPLDIK--EMPMTIGEGTLP 360
Db 262 ISHPDSG-----ALMIYDSAAACELAVSAYAMLEASHHTFASLIYNDWPIPLADCEGHA 315
QY 361 RLNLIGSTAVAGYIGAMVFPNSALYLTLEVEDSGMTAAKDGGPSPFNFPYQFAG-- 417
Db 316 RYVALHRYNASLIAPVSTOIFATNSVLY-----SGVSKSTGGQKESLEFNFYMTIHGIGT 370
QY 418 -----PHLAANQTDQRDGVIVSSQSTGSSNTFFSVDTALICGPAPILARILP 466
Db 371 LQETWDPCKRRPCSGMGGDPVTG-----TNGPQIN--YAVEHLVYVASFNPILLARAY 422

```



```

OY 356 EGTPLRLNALGSTARAVAGVIGAMVSPNSALTYLEVEDSGMTEAKDGGPSENFYOF 415
DB 315 ETEEDRIKAIQIMNAMSVHYTHLPSTNSVLYLTKINQOTOSKSEON---YNYTYFMQ 371
OY 416 AGPHIAANQOTDRDGVIVSSQSSNTESVYLYALICGFGAPLLARLLFYLERCDAGA 475
DB 372 KDLVAADATQURENDEBPASGAFKESGGTYTLTHLALASSFPHLLAKNYYMQFCVHOK 431
OY 476 FTGGCHDAIKYVYGFDSHPCSLCEKHTKRPVCAHTVHRQRMDR-PCQATROPICGV 534
DB 432 STTNANYSVPQVYTAASADICLCOGTCRASCITHTLFLKLRPRPVLCOSQRDPVYV 491
OY 535 GTMNSQYSDCDPLGNAPYLLLRKPGDQTEAAKATMODTYRATLERLFTLEDEPLDPA 594
DB 492 G-VSGQYNDLDMIGNAFTE---KEKDEAVQNAESEKTYTMOIQWV---EKLSTMG 542
OY 595 ACCSSGGLSSVVDHPTFRHLLDTLAKILEOTTPQPMKVIYERDOKIRGGLSEATHSMA 654
DB 543 VTEGTVG-SELTIDIOSLTKTRFDIDNVDSYVKFMCLVK-NINLNETTKTVHVLH 600
OY 655 LTFDPYSGAFCDITNPLVNRHLAVVQDLAL-----SQCHCFYGOOV 697
DB 601 YCCNVEWQAPCAMPLNIIFYKSVLAIIDICLPLAMTYEQDNPSICMMPSEMLKVH--QT 658
OY 698 EGRNR-----NQPQVLRKRFVDLNGGISTRTITVITSEGPVSAFNPTLQ 746
DB 659 IWNFKSSCLDRCQVLTGSHKIVHTDMPDPLN-----IDSALS-----CQ 699
OY 747 DAPAGHTFUGDLAVHSEVYIRDIRVKNRVVFGNCTNLSEAAARLVGLASAYQROEKV 806
DB 700 IVP--MKQVRLAKALLTPKTIKIKRIVFSN--SSMETTIOSGIKSAT-----KKS 750
OY 807 DMJHAIQILKOPHGLLPYKGMPSKSPNPQWFTLLQNRQMPA-DKLTHEITIA 864
DB 751 YLVTPYKMFINSKHKVMP-----PNAKISALYLMHTFESQKQOLPVLPCISREHVELAN 805
OY 865 AVKRETEEAALNFINDPCTCIGELAQFYMANLILKYDHSQYLINTLSI-----ITG 918
DB 806 YVERSSKMHDMNVADILPTTLITLAKVLANNTILKTCGOTQVATTLQCLPTLQFTISA 865
OY 919 AKRPR--DPS--SVLHWIK-KD-----VTSADIEQAQKALLEKTENLPELMTTAF 964
DB 866 TEYPIVLLDOSIMSDHYLSSIKDKHALTVQTLKEDATVK-----908
OY 965 ISTHLYKAAANRPVAVICISISKYHGAAGNKRVOAGN---WSGINGCKNVCPLFT-KD 1020
DB 909 -----QRPVTVPLVANKYTGINGNQIFQCCNIGYFMGCGVDNRNLPDSTGR 957
OY 1021 FTRFIIACPRGFTCP---VTGSSSGNRET---TLDQVRGIIIVSGAMVOLAIVA 1071
DB 958 KQNNSSYKRRHNVHFMVMAHLVKKNSNINNTPEVETIRKRVONIFELKD---NINIFD 1014
OY 1072 TV---VAVVGARQAHMAHDMLSITDPELARIPELHNOITITQIETPWTEGALFAV 1126
DB 1015 NVLELVKGLDSCENITEDDQFYLGYYIMSDEIWSRFQIITDSGAWSEVENTKVL 1073

```

```

RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Estathiou S., Craxton M., Mccallay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RT Virology 209:29-51(1995).
CC -1- FUNCTION: SINGLE STRAND DNA-BINDING PROTEIN REQUITES FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL databank.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: X92436; CAA63167.1; -
DR EMBL: X83413; CAA58375.1; -
DR InterPro: IPR000635; Viral_DNA_bind
DR Pfam: PF00747; Viral_DNA_bp.1.
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN-FING 459 475
SQ SEQUENCE 1132 AA; 127762 MW; 4468D3E559F02D8 CRC64;

Query Match 7.8%; Score 492.5; Pos 1; Length 1132;
Best Local Similarity 21.8%; Pred. No. 1,2e-29;
Matches 274; Conservative 185; Mismatches 544; Indels 253; Gaps 53;

OY 2 ENTQKTVPTGPGYVACRVEDLDEISFLAASDSDLLPLMRILYKTEKTS 61
DB 3 DENEIVSAPVSTAAMLTYPPKDELDVLSLBERNSPIYSLMLTYHNDPSTT 62
OY 62 LAVVSARITGLAGAGITLKTSHFPYVYVYHSGKHLV-PSSAANLTRACNAAPERF 120
DB 63 V-----KIPITNFGIILTKITSEMPVCFHFJTEGLVGMEDHDLIRLEQTRUKF 115
OY 121 GFSRCQPPVDGAVETGAEICRLECPENTILYVAVTALFEAVPMCVPLHYGID 180
DB 116 HUSFEVPTARKYIDIKALCSAVGKADSVYCHVACNGFEELF-----AGLLI 165
OY 181 -----VHINGVVIRIPLPVQLEMPDVNLPVDPENTHRSIGEGVYPTPYNTG 232
DB 166 PCVEEQIQVOVGYSVCVILPYSATILETE-----ETISLSSOTFLOKRCPTIA 216
OY 233 LCHLHNDVYAPRAVALRPNVTAVAAPGAHILAFENHFPNAVLPDITTYFOSSSGTT 292
DB 217 LSETLTYVFTSKGTLIRSNTELIDAGLKQFTQDGEYLVKLAPKTYL-----266
OY 293 TARGARNRVNSTSKPSPGCFERRIASIMADPTALHAEVFN-----TGVEFTPTDI 346
DB 267 ---GISQKISAVEK-----FLMVDQV-TELSFHVAYELISVVD--PSQI 309
OY 347 ---KEMPMFIGMEGTLPRLNALGSTARAVAGVIGAMVSPNSALTYLEVE-----D 394
DB 310 MNFNDMPILIRNSETTAERMAOLTNLKLHLSHLAVLIFAPNSILYOSKLAFLIPNVQAFN 369
OY 395 SMTTEAKDGAPGSENFYGFAGPHLAAMPQTRDGH-VLSSQSTSSNTESVVDLALI 453
DB 370 SWMTQ-----ELLRLSIFCNALSSLDYVNNRKKIITKISTGSKDKFSAHLAVA 422
OY 454 CGFGAPLLARLLFYLERCDAGAFGCHGALKYVYGFDPSEIPCSLCEKHTRPVCAHTTV 513
DB 423 CATSPQLSFEVYVNNLRMSVYNAHNAHTEIYHNLVNC--SANLCEFGDCKCQSCIGTAM 480
OY 514 HRLQMPREFGATROPICVFGTMNSQYSDCDPLGNAPYLLLRKPGDQTEAAKATMODT 573
DB 481 VAVGTRLPALIPKNNKKEPLVMSKFSRYTAEVDLGSFG-----RKPVSLEKELGKDOQNI 535
OY 574 YRATLERFLDEQRLDRCA-----PCSSGGLSSVVDHPTFR-----RIIDT 618

```

```

146 546 LS LDREKVSJIFVYKKNSLIFVTSFETFNVSCKKLEVSIII 580
147 619 LKARLEQTTOFMKVIVLRYKI REGISSEATHSMALTFDYSACPGITNPLVKRTHL 677
148 581 LFCVTEKVS KQIVKRRKUTTRBUJENCLQSNVITTYATAFSPHLYSYKVL 637
149 678 AVVDALASQCHVYVGVGVFGRNFPNPGVPIPEPVDF NAGFISTRIT 729
150 648 LVLQMLAL IVASGHVVDKPCIGNSIKMLVOYQSLYCHFHSSYLAKKPLNRIYK 693
151 740 VTLSEPPASAPNPLDQ DA PAGR FPGDILARVAVVRIQVKNRVVPSGN 780
152 694 V ASVNDMEQLDLDLYKSGKVKTTGAKLRLSMGCLROGRINPKNSS 745
153 781 CTNSSEARARIVGLASAVQROEKRVDMHGAIGLKKOFGHGLFPRGMPKSPNK 839
154 746 KFAHNPPYFKKV KHAKNLSQSTSLKPKYHKKL PAVKISGLFL 791
155 840 MEWTLIDRQMPAKLTIEHTT IAAVKRTEEYVAINEFNLPTGIGELAOFYMANLI 898
156 742 WQRELLNNVPKLIDQNDPEVKTFIKFASTINIDEIDIDIDPECLSTFIDCYFHKKF 851
157 899 LKVTDSOYL INLISLITG ARPRDPS SVLHMIKQVTSADILET 944
158 852 LSALEFHDTLSLRKLSKLVTONVLEPVLDKQPKSSIDCYLVYKVLDDG VBN 909
159 945 QAKALEKTEINPELMTAFSTHVRAMNORPVVGLSISKYGAAGNRRVFOAG 1002
160 910 FVLSLSKPN FGLHFS KSLVFLGLLEKFLVSLA NMEYFOFQOL 954
161 1003 NMSGLNG GKKNCP FTEDTRGFLIACPRGGITCYVTGPPSSNRETISDQ 1053
162 955 CMIGSGVGRNINPLSSALQDPRMK QTLTATKPSVIV KVKYKRLAIMPTRV 1008
163 1054 VAGTIVS GQAMVQIATVAVRVARGAQOHMAFDMLSLDD EFLARD 1102
164 1009 VKGKVLSTENTNDIDPELLIAEVM DREDKPTMDQLFFVDGSEALASIMKL 1065
165 1103 KEMHOUTTLETFWFGALPAAV KILDEKTTAGDC EPTT 1142
166 1066 NMLVIMNKKPSEI NQSPFVSSNDAPVYDFSELLAEEDQNGVAKCDELET 1120

RESULT 13
167 DMBI HSV67 STANDARD PROT 1132 AA
168 52548
169 01 OCT 1996 (Rel. 34, Created)
170 01 OCT 1996 (Rel. 34, Last sequence update)
171 16 OCT 2001 (Rel. 40, Last annotation update)
172 Major DNA binding protein (MMP).
173 541 OK KAL.
174 Human herpesvirus (type 6 / strain 729) (HHV6).
175 Viruses, dsDNA viruses, no RNA stage, Herpesviridae.
176 Herpesviridae: Kaposi's sarcoma-associated herpesvirus.
177 NCBI TaxID:36351.
178 111
179 SEQUENCE FROM N. A.
180 MEDLINE 95074921; PubMed-7984761;
181 Stanley F.R., Damjanovic G., Black J.B., Damhaugh T.R., Pellet P.E.;
182 "Molecular cloning and amplification of human herpesvirus 6B orf1y1
183 suggests acquisition of orf1y1 by transposition";
184 J. Virol. 69:589-596(1995).
185 1. FUNCTION: SINGLE STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
186 REPLICATION.
187 1. SUBCELLULAR LOCATION: Nuclear (Probable).
188 1. SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
189 FAMILY.
190
191 This SwissProt entry is copyright. It is produced through a collaboration
192 between the Swiss Institute of Bioinformatics and the EMBL Outstation
193 at the European Bioinformatics Institute. There are no restrictions on its
194 use by non-profit institutions as long as its content is in no way

```

```

195 modified and this statement is not removed. Image by and for commercial
196 entities requires a license agreement. (See http://www.isb.scrib.announce/
197 or send an email to license@isb.scrib.ch.)
198
199 DR EMBL: AF157706; AAB06339.1;
200 DR InterPro: IPR000635; Viral_DNA_bind.
201 DR Pfam: PF00747; Viral_DNA_bp.1.
202 DR DNA-binding: DNA replication; Zinc-finger; Nuclear protein.
203 FT ZN.FING 459 475
204 SO SEQUENCE 1132 AA; 127817 MW; 839AB46A5104HP CRC64;
205
206 Query Match 7.8%; Score 490.5; DB 1; Length 1132;
207 Best Local Similarity 22.0%; Pred. No. 1.7e-29;
208 Matches 278; Conservative 180; Mismatches 535; Indels 271; Gaps 54;
209
210 2 ENIGKIVYPTGPIGVACVCHVDLLEHLSPLAASTSDIALPLMNLVYKIFTS 61
211 3 DENETVASPVSTAMVYFPEKEKELDLVSLSTMERSPVVISPLMLNLTVENDEFT 62
212 62 LAVVSGARTGLAGSITLKLTHSFPVFEVFGKRYL PSHAPNLTPLNMAREF 120
213 63 V KTLPTNGGTLITKITSFMYCPFHGTRQIVCAHHDGDLIRICRQKRF 115
214 121 GFSKCGPVDGAVETGAEICTRGLPEENTILVLTALKEAVYKCNVFLHYGDI 180
215 116 HUSEEVPFAFVVIDKA LSAVGRKADSVIHWAGNGSEKELLE AGLLI 165
216 181 VHIHGVIRIPVVOIFMPDVNKLVPDPTNHRSGRQVYPTPYNG 232
217 166 PVEEDYOVGEYSVKPLYSATIFETE ETLSLSTGTFIDRGLFLA 216
218 233 LGHLIDCVIADMAVALPVNVTAVAPGAHLAFDENHGAVALPDITVYFGSSSGTT 292
219 217 LSETLFFVYVFTSMGTLIRSNTEKELIDAGIKOFTQGGQVKAIRKIVL 266
220 293 TARGARNDVNSTKSPSGRFERPLASIMADPTALHAIVIN TGIYEETPTDI 346
221 267 GISGOKISAVEKO FLMLVDSPV TELSSHVAEYLDSDYD PSOI 309
222 347 KMPPTGMRGCTIPRINALGSYARVAGVIGAMVSPNSALVLEF 394
223 310 MNENMPITIRNSETHAERMAQTLNKLHSHLAVLIFAPNSILYCKSLAFIPNVQAN 369
224 395 SGMTEAKDGSGPSFKRFQVAGPHLAAMPQIDKQH VLSQSTSSSTERSVDLALI 453
225 370 SVMTQ ELILRSLSFENALSSLPEDVYNDNRKLIKDSHSGDKDKESANHLAYA 422
226 454 CGEGAPILARLLEFLERCAGAFGTGCHPDALKYVTGFSEI PSLFERH 503
227 423 CATGGLLSYVVMNLMKSV YKGNATELYNHLVNSANLCEFTDGR 470
228 504 TRPVCAHTTVHRLKQMPHFGATIKQPIGVPTMNSOVSDDPHGNVAVYLILKQDQDT 563
229 471 CQSGTGTAMVAVGTPPLPAIPKRVKREPLVMSNFSYVAFVITNSFG FKEVSEL 525
230 564 EAKATMODTYRATLERLFDLEQELLDRGA PSESELSSVYDHPTR 613
231 526 KETGKQDQNTLS LDRKPVSDILOYCKKNSLIDVVTGSDIIVNVS 570
232 614 PILDTEPARIEQTTOFMKVIVLRYKI REGISSEATHSMALTFDYSACPGI 667
233 571 KKDPSILHGLQIEHVS KQIVKRRKUTTRBUJENCLQSNVITTYATAFSPH 627
234 668 TNPVAKRTHLAVQDLASQCHVYVGVGVFGRNFPNPGVPIPEPVDF NAGFISTRIT 729
235 628 LTFAYYKVLTLQMLAL IVASGHVVDKPCIGNSIKMLVOYQSLYCHFHSSYLK 683
236 720 GGFISTRITVLSKSPVAPNPLDQ DA PAGR FPGDILARVAVVRIQVKNRV 770
237 684 KQIVNRIYK ASVNDMEQLDLDLYKSGKVKTTGAKLRLSMGCLROGRINPKN 735
238 771 VKNRVVSGCNTLSSEMAPRLVGLASAVQROEKRVDMHGAIGLKKOFGHGLFPRGMP 830

```

```

Db 736 IKRPNKSSKAHNNPDKKNV-----KHKKNPLSGCISFLPKYHDKLF----- 781
QY 831 PNRK-SPNQWMTLLQBNAMPADKLTHERIT-IAAVKFTEEYAAININLPTCIGE 888
Db 782 PNRKSCLEHMOFELNNPKLIDIGNPEVKFKFIFASITNTYDEIDIDIQPELST 841
QY 889 LAQFVMANLILKYCDHSDYI--INTLTS-1ITG-----ARRPRDS--SVLHWIRK 934
Db 842 FIDCYHNNKFLASGHFDLITSLHGLTSKLYTONPVPFVYLDKOKRFFSIQIEYLYVK 901
QY 935 DVTSAADITQAKALLEKTEENPELMTAFTSHLVRAANRPVYLGISISKYHGAAG 994
Db 902 LVLDG--VNPVNPVLAISKPN--FCITPTS-----RSLVTFGLTLKFPVSLA- 944
QY 995 NNRVFOAG--NMSGNG-KGNVCPPL-----FTEDFRRIITACPPSGGFCPTGSSGN 1045
Db 945 NNEYFOGOLGWTGGSGVDRNLNPTSSALODEFMR-QKTIITATKSEVIV-----KVR 998
QY 1046 KETTSISQ--VWQIIVS-----GQAMVCLAIYAIVKAVGARAQHMADMLSLTDD-F 1096
Db 999 RETIMEDTEVYKGVLSIVENTLNDIDPELLITAEVVR--DREDKPTMDMLFVYDRE 1055
QY 1097 FLARD-----LEELHDDIQTLETPWVEGLEAV-----KILDEKTAGDS--- 1138
Db 1056 ALAASIMLKINHLDVNNVDPFSA-NIQSVFAVSSNDAPVYDFSHLLAEEDDQASGLK 1114
QY 1139 --ETPT 1142
Db 1115 CDETER 1120

RESULT 14
ID DBI_HSV7J STANDARD: PRT; 1131 AA.
AC P52339;
DT 01-OCT-1996 (rel. 34. Created)
DT 01-OCT-1996 (rel. 34. Last sequence update)
DT 16-OCT-2001 (rel. 40. Last annotation update)
DE Major DNA-binding protein (MDRP)
GN 041.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=57278;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J1;
RA Nicholas J.;
RT Submitted (JAN-1996) to the EMBL/Genbank/DDJ databases.
CC -! FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
    REPLICATION.
CC -! SOURCE/CELLULAR LOCATION: Nuclear (Probable).
CC -! SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
    FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U43400; AAC54703;
CC
CC InterPro: IPR000635; Viral_DNA_bind.
CC
CC Pfam: PF00747; Viral_DNA_bp; 1.
CC
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC
CC ZN_FING 459 475 C2HC-TYPE.
CC
CC SEQUENCE 1131 AA; 129008 MW; 52C97388D1B6D04F CRC64;
CC
Query Match 7.7%; Score 483; DB 1; Length 1131;
Best local Similarity 20.8%; Pred. No. 6,4e-29;
Matches 258; Conservative 204; Mismatches 560; Indels 220; Gaps 46;

```

```

QY 2 ENQKTVTPPGUGVYACRAVEDLDLEISFLAARSTDSIALPLMRNLTVKRTTSS 61
Db 3 DDETVVASICTAAWLLYLPKEOKLEILTLLSLMEKRKSVYSPLLNLTVENDEFFT 62
QY 62 LAVVGARTTGAGAGITLTKTSHFPVSFVFPGGKHVLFSSA-ANLPHACAAAPRP 120
Db 63 V-----KPIILNGTIVITKITSFMPVCFPPFGTIVPLKFAADPHONLKLCKQTEKK 115
QY 121 GFSRCGGPVYDGAVENTTG-AEITRGLPEPENTILYLVYALFEAVFMCNVFLHYGGL 179
Db 116 NL--QEFVNVGNMKSVDICKICHSVGKNADVLCHIVGNGFELLF-----AGLL 164
QY 180 I-----VHLNHDP--VIRIPLPVOLFEMDPVNNLVDPPTNHRSLIGEGVYTPPYNT 231
Db 165 IPCVEEQIQVQVEGCAIKIPLYSATLFESEETICT-----DTCTEPIQENGYAP 215
QY 232 GLCHLHDCVIAPMVAVALRVNRVYAVARGAHLAFDENHRCGAVLPDITVYFQSSSGT 291
Db 216 QISEVLFYLLFTSMGMLRKNNTLELIKAGLKQPIQUDPQVLAIRKTY----- 265
QY 292 TTARGARNDVNSTSKRSPSGGFEERLASTMADTLHAETPN-----TGIEETP-T 344
Db 266 --HGIPGQKLSPIEKD-----HMLVD-AVITETITSYAEYLDIYENNOIM 310
QY 345 DIKEMPMFICMEGLPRLNALGSTARVAGVIGAMVSPNSALYLTVEEDSGMTEAKDCG 404
Db 311 NFSEWPIKSAETHEEKIVELKRLHLSHVAALVAANSILYSNLAITYSNK----- 365
QY 405 PGPSFN-----RFPQAGPHLAANPQTDRODH-VLSQSTGSSNTPSVYDIALIC 454
Db 366 --QAFNSAITQETLIRNSQNSLSENEDEFYNDARKLIKNSSPKCKEDKSAFHAYAG 423
QY 455 GFGAPILARLLFLEKRDCAAFITGSHDALKYVGTGDSLPGLCKKTHRPVANHITVH 514
Db 424 ATCPQIILSHIIMNINRISITNTGNSIYNHIVNC--SSNLCEFCGCKCHSCITGAL 481
QY 515 RLKORMPRQOATROPVGVGTNNQSYSDCDPLQNPAPYLIIRKPGQUTAAKATMDQTY 574
Db 482 RINSRLPQISKTTRKEETVMTESRFYADVGLSGFKKGVNEKDKMEQOTTPSLDRF 541
QY 575 KATLERFLDLBOBRILDRAPCSSGGLSSYIV--DHPTTR-----RIIDTLARAKI 626
Db 542 K-----FLGMIDY-----CKKNLIDAIIGEDMINPKSUNDVNMINDLIQCLIEA 588
QY 627 TTQFMKVLVETRDYKIREGSEAIHSMALTEDDPSAFCPTNLNLYKRIHLAVQDALS 686
Db 589 VSKCISMRTQ--TSRQIENCLQSFNIDTPLSLAFSPFPVYTYKVALIVLQNLAL- 645
QY 687 QCHGVFYGOQVE-----GRNFRNOFQVLLRRRFDVLFNGCFITRSITVYLSFGPVS 738
Db 646 ---IIGTVYVDDPCTGNLISKMLMOQYQSLGAFYNSHFKKGLNKTKYKI----- 694
QY 739 APNPITLQ-----DAVAGRTPDGDLARVSEVYRDIIRVKNRVVYFSGNCTLSAAR 789
Db 695 ASNVDMQYIDPMLFSGKTAKTISIAKLCRLSMQCLDKRVKRRPNKRNKNTQNN- 751
QY 790 ARLVGLASAYQREKRYDMLGALGFLKQFHGLFPRGMPNPKSPSPNQWMTLLQBNQ 849
Db 752 -----PFFKVKYQKKNPGLSGCISFLPKYHERLF-----PNLKISCLR-FMQRIILNN 798
QY 850 MP--ADKLTHEEITT IAAVKRTEYAAININLPTCIGELAQFMANLILKYCHSV 906
Db 799 MPKTIIDGNVELMKRSFKETFKVYNSVDEIDLDIQECLLSFIEYFHNKLSVLYKRD 858
QY 907 YL--INTIISIIGARRPRDPSVLIHWIRKDY--SAAUJETQAKALLEKTEENPELMT 962
Db 859 YLNSLAHLISKVL---PQNPMLFPVFLKRLKHPFTSSQGVQEVYMHAKVIV--GNGIKLPMYA 912
QY 963 AFTSTHLYRAANRPVYLVGISISKYHGAAGNN--RVFOAGNKGSGINGKKNVCPPL- 1016
Db 913 SLTRKPNFGSIFTRGSLIITPGMLIEKVFVSAVRDYEFHFGQIGWLAGSGVIRNINPSPSGI 972

```

```

01 1017 FTHDGRKRFELTAVYRGGELTGVGSSGNETLTSD--GVKGLI-----VSCGAMV 1066
02      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
03 9/3 GQFGRMR-QKEVIALR-----LGDII-VKKYKREAIYVEVETRCGLVNIETLSNSVNE 1026
04      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
05 1067 LAIVTVYAVAVARAGQAAMFDMWLSTDD-ELFARLDE-----ELHDIOTLTETP 1116
06      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
07 1027 LLLIAEVMK--DRDSKPLMDMLPYVDORFLAKSYMNRKIGLITDILNVHDFSLSTLSTV 1083
08      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
09 1117 W-----TVEALEAVKIIDERTAGDDEPT 1142
10      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
11 1004 FFEQVEDSNAIYQISELLVEVNEQSGFGLCKELETENEERS 1125
12      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
13
14 RESULT 15
15 ENH1 SEQVC
16 ENH1 SEQVC STANDARD; PRT: 1160 AA.
17 P14215;
18 01 JAN 1990 (Ref. 13, Created)
19 01 APR 1993 (Ref. 25, Last sequence update)
20 16 OCT 2001 (Ref. 40, Last annotation update)
21 Major DNA binding protein (MDBP).
22 1167 OR DBP.
23 Simian cytomagalovirus (strain Colburn).
24 Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
25 Herpesvirinae; Cytomegalovirus.
26 WHI Taxid:50292;
27 (11)
28 SEQUENCE FROM N.A.
29 MEDLINE:91037979; PubMed-2172458;
30 Anders D.C.;
31 "Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding
32 protein gene: comparison with alpha and gammaherpesvirus
33 counterparts reveals conserved segments."
34 J. Gen. Virol. 71:2451-2456(1990).
35 (12)
36 SEQUENCE OF 205-308 FROM N.A.
37 MEDLINE:88155776; PubMed-283198;
38 Anders D.C.; Gibson W.;
39 "Location, transcript analysis, and partial nucleotide sequence of
40 the cytomegalovirus gene encoding an early DNA-binding protein with
41 similarities to ICP8 of herpes simplex virus type 1."
42 J. Virol. 62:1464-1472(1988).
43 1. FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN (REQUIRED FOR DNA
44 REPLICATION.
45 1. SUBCELLULAR LOCATION: Nuclear (Probable).
46 1. SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
47 FAMILY.
48
49 THIS SWISS PROT entry is copyright. It is produced through a collaboration
50 between the Swiss Institute of Bioinformatics and the EMBL outstation at
51 the European Bioinformatics Institute. There are no restrictions on its
52 use by non-profit institutions as long as its content is in no way
53 modified and this statement is not removed. Usage by and for commercial
54 entities requires a license agreement (See http://www.isb.ch/announce/
55 or send an email to license@isb.slb.ch).
56
57 EMBL: U008750; BAA00647.1;
58 EMBL: M19868; AAA46066.1; ALT_SEQ.
59 PIR: A66256; A66256.
60 InterPro: IPR000655; Viral_DNA-bind.
61 Pfam: PF00747; Viral_DNA_bp_1.
62 DNA binding; DNA replication; Zinc finger; Nuclear protein;
63 Early protein.
64 ZN FINE: 465 479 C2HC TYPE.
65 SEQUENCE 1160 AA: 129005 MW: 70477161616974BIA CRG64;
66
67 Query Match 7.6%; Score 479.5; DB 1; Length 1160;
68 Best Local Similarity 21.3%; Pred. No. 1,26,28;
69 Matches 279; Conservative 202; Mismatches 571; Indels 255; Gaps 53;
70
71 1 MENTUQKTVTVTGLPGVAYACVVELDLDEISFLAARSTRDNLALLPLMRNLTVKRTTS 60
72 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
73 1 MSNRLSALAVGVAAVYVFTKTNHNNVIALISLQSSSVVIALPLIMGLITVDDPCT 60
74 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

0Y	61	SLAAVSCARTTGLIAGAGITLKLITLTHSHYPSVPHQCKHVLDP--SSAPNITLTAACNAARH	111
Db	61	SV-----RTVVCYDGSVGLIKTSSFCPPALTFYNTQGVDSBPHGVQLDETQJR	113
0Y	120	PGFSQCGCPVDAQVAVETTGAFICTHGLLEPENTLIVLVTALPKFAVCMCNVFIHYCGID	179
Db	114	VAIESYM--PREGAPRPIDLAICATAACGDDQJVLVHVVGNCMKRFFMYAGOLIPCEHFAA	171
0Y	180	IVHMGVGVITPLPEPOLF--MPDVNRLVDPENPTHHSIGEGVYVPTPEYNTGTLNL	246
Db	172	PTIRINQIAVAVVPIYPTPTGSLQAVDD---SHELILNKSS--SVWISRCIVVAVSPT	225
0Y	237	IHDCVIAEVAVALVVRKVVATAVARCAAHAPDENHGAQVLEPDDITTYTHPOSSSGTITARG	296
Db	226	LEFYVYTSMCQALPFSFTKVLIPALIKLQFVNDSDQVCKLAPKKHYGTYSQK	277
0Y	297	ARRNVNSTSKPSISGCHERLIASIMADITL-----HAEVIFNGICIEETPIDIKMP	450
Db	278	-----ISSLENDP-----HMLSDIAVEICELGSPASVYLDASVQASDSMYSTMP	321
0Y	351	MEIGNEGTLPLPLNMAISYTAARAVAVIYAAVFEPSNAL-----YLTEVDSQMTFAKRG	404
Db	322	VYVNAIDHROHILIRALFELIKLHLSHISALLPGCSNLSLHNRLVYLSSNNAGST---GA	377
0Y	405	PGPSPNREYQACHILAAINQTDHCH--VLSOSQICSSNTHPSVDYALIGCHGAPILAR	463
Db	378	SOEVLKSIHFANGLTGLCEITYNDAKLLIKGQGVAKDERAPYILSLIGTCFOLPSA	437
0Y	464	ILFYLEKRDACAFIGCHQDALKYVTCFDEP-----IPC--SLICK-----HIRVCA	509
Db	438	FILWYLNKSV-----YNTGLGSSITLSNHLIGCSSSLCAGACGCTCH---CY	482
0Y	510	HTVHRLQRMPREFQATKQPIGVFGTMSQVSDCLPLNVAEYLLKRPQJGLEA---	565
Db	483	NTAEVAVQTRIDPQMPRIPKKEPSVAVQSRFLINDVYLTCG---RYSASRCKEASIDA	538
0Y	566	-----AKAIMQDTPATILELPIFDLQERLLIDRCAPGSSRGISSVYD--HPT	611
Db	539	KADEGSASTSKRTASSSVDRHR--LNRLLDYCKKMPFLID-----SVTEEDMTINGSD	591
0Y	612	FRKILDTLKAIRFOPTIQFMKVLVEFRDVKIRGCLSEATHSMALTFPDYSCAFCTNHL	671
Db	592	FLINIYSSLNKTVDEAMSPVS--EYHMKSNRDEVLGATQANLIDINPAVSPSLAYE	648
0Y	672	VKPTHLAVVGLALS--HVFVYQ-----LVEIRFPFN--PVLPLRFRFVLLNCHLSTPS	727
Db	649	YRVRIFALIGVVALITAVSYIVDNPITTSLSVSWVYQHOSI-----HGASTYSS	699
0Y	728	ITVTLSKCPVAPRPDTLGGAPACR-----TFQGLIARVSVFIRIDIVK	772
Db	700	-----RKGFLEIRVWCKSSKANDIRLPDQFLYARGYTVSISMEIKLSPSVSLMFEYK	754
0Y	773	NRVYFSGNCITNLSPAARARLVGLASAVORQ--KRVIMLGAIGLILKQPHQILPRQMP	830
Db	755	NRPI-----SKASKGT---TAHVFFRHHVKKYNPKCIGLPILYKHDKLFP---	799
0Y	831	PNKSPNPQWMTLLOJONOPADKLTHETITLAAVCRF-----TEEYALINFINLP	883
Db	800	-----IDGFSQSLQFMQVCANALP-----KNVNLDJMGCPNNVVKFVLSVLAJYVNHDI	853
0Y	884	TGICHLAQPYAMANLILKYCHSGQV---INTNLSLITGARRPRDPSVLAHMKLIVASAD	941
Db	854	DCMLNYLENPHNKKFCFVYFQKDYITGLTHLTLPLLYQNHALFP---YLIGAPNPAVSDA	911
0Y	942	LETOAKALLEKTENILPLMLTAAFLSTCHVHAANNORPMVVLIGISISKYHCAAGNRRVQA	1001
Db	912	FALRIKLDEL--KATGVAPLAASTVIRSLIMRTLPEQSLTVASHSLEKAVGVANNKEILYQF	969
0Y	1002	GN-----MSG-----LN-----GSKNNVPLTFPTPTREFITLACPRGSGICVGTSPSSGNET	1048
Db	970	GOIGYFSGNVGSLNINSSIGCGI-----YAFKFMVQCILAKUSIVL1-----KSRSDN	1019

```

QY 1049 TISDQVKG11VSCGAMVOL-----A1YATVRAVAGARAOHMAFDMLSLTDD- 1095
Db 1020 VLFDE---DIKRVMAALDSENLDYDPELMAMF-----ETLSTREEIPERDDVLEFVDGC 1072
QY 1096 EFLARDLEELHDOIIQTLTPWTVEGALEAVKILDEKTTAGDETPNTIAFNFDSCPSH 1155
Db 1073 QAVADSLMEKFSRIOPMGVDDFSL---VNLOQVILDSRPECCEGCGEV-----H 1117
QY 1156 DTSNVLNIGSNISSGTVGKL-RPPED--DELFDSGIPIKHGN 1199
Db 1118 DLSALFTAASGEAVGNSV--GLNARGGEHAFDE--DGLLPARKGR 1160

```

Search completed: March 28, 2003, 13:36:03
 Job time : 37 secs

